Department Of Mathematics

Statistical Methodology for Tab-Charts : Data Reduction Techniques in Laser Ablation Analyses

by

YUNG, Chi Ho, BA(Hons), MSc, MA

Submitted in fulfilment of the requirements for the Degree of Master of Science

University of Tasmania (Sept, 2008)

Acknowledgements

I am deeply and greatly indebted to my supervisors Dr Des FitzGerald and Dr Peter McGoldrick for their valuable advice and encouragement. I am also thankful to those who gave me their useful help in my course.

Authority Of Access

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

YUNG, Chi-ho

Date : 15th September 2008

Declaration

Except as stated within this thesis, it contains no material which has been accepted for the award of any degree in any university. To the best of my knowledge, this thesis contains no material previously published or written by other people, except where due reference is made in the text of this thesis.

YUNG, Chi-ho

Date : 15th September 2008

Thesis Abstract

The researchers in the CODES and the School of Earth Sciences operate a laboratory to study the composition of rock samples, which are collected from the field site. The rock samples are put into a machine. This machine will create series plots of all elements (called tab-chart), which indicate the distribution of elements in the samples. In the tab-chart, a significant signal change implies the change of composition in the sample and a flat part implies a mineral layer (phase) existing in the sample. Currently, the researchers identify these properties by their knowledge and experience. In some situations, they are difficult to make their judgement on these properties since they are not obvious and clear. Thus, an automatic and systematic method is requested to help them to solve this problem.

Total 1848 (= 66 samples × 28 elements in each sample) tab-charts of primary and secondary samples are provided by the School of Earth Sciences. These primary and secondary samples are not real and created in the laboratory. These tab-charts have the shape of background noise (a flat part) at the first stage, jump (a significant signal change) at the second stage, plateau (another flat part) at the third stage and drop (another significant signal change) at the last stage. Although this project focus on the standard samples only, the analysis and results can be extended to the real samples. The first four chapters of this project explain and describe the equipments of the laboratory, the mechanism and process of geological analysis on the sample and tab-chart description. These chapters provide the knowledge for reference only and not the main interest in this project. This project is to focus and concentrate on the mathematical analysis on the tab-chart.

The problem mentioned in the first paragraph is actually a change point analysis (detection) or time series segmentation issue in mathematics and statistics. Many methods are presented and invented to solve this problem in the papers. They include cumulative sums of difference (CUSUM), perceptually important points (PIP), fuzzy set theory and genetic algorithm and so on. To my best knowledge, most of these methods focus on point change detection only. On the top of this detection, a method is expected that it can also provide the researchers about the statistical summaries of the flat part in the tab-chart (i.e. the mean, standard deviation and trend of element amount in the layer). Therefore, time series model could be considered and a good choice to achieve the above two targets. In addition, time series model has the advantage that it is easily implemented in the worksheet. However, some algorithms and modifications are needed to make such that the time series model can identify any point change in the tab-chart.

Among various time series models, the linear Holt exponential smoothing model is selected in this project. This selection is made after considering and comparing some common and popular time series models. For simple exponential smoothing model, it has one estimate or equation (i.e. smooth) and one parameter (i.e. α) only. This model is rejected because signal changes (i.e. jump and drop)

have large slope but flat parts (i.e. background and plateau) have gentle slope. One estimate is not enough to reflect this slope property of the tab-chart. For damped-trend linear exponential smoothing model, it has two estimates or equations (i.e. smooth and trend) and three parameters (i.e. α and β and γ). Although two estimates are enough to reflect the slope property, three parameters may complicate the problem analysis and there is another better choice, linear Holt exponential smoothing model. This model also has two estimates (two equations) but has two parameters only (i.e. α and β). It is not guarantee that this model is the best choice, but any results and findings from this model can help to explore other time series model in the further studies.

The linear exponential smoothing model is modified before trying to fit it to the tab-chart. The modified model has the variable (dynamic) parameters, α and β, and a threshold value, T. In the fitting process, if the trend estimate of the model exceeds the threshold value, the variable parameters will take values $\alpha 1$ and $\beta 1$. Otherwise, they will take another values $\alpha 2$ and $\beta 2$. The reason of using this policy is based on the difference of slope between significant signal changes (i.e. background and plateau) and flat parts (i.e. jump and drop). The parameters and the threshold value are adjusted manually until the model is fitted well to the tab-chart. After fitting the model to all tab-charts, it discovered that the threshold value T is more influential and important in finding the well-fitted model than the two parameters α and β. Besides, the fitted curve of the model is spiky when the threshold value is small but becomes smooth when the value gets larger. There is a remark that the above parameters policy is only an initial trial and not perfect, the experiment result will reflect and reveal what is the drawback and disadvantage.

For convenience, henceforth HOLT model is named for the above modified linear smoothing model. The first algorithm of detecting the point change (or time series segmentation) by HOLT model is called classification method. The main idea of the algorithm is explained at the following. The change of the variable parameters (i.e. α and β) indicates the stage change in the tab-chart (i.e. from significant signal change to flat part or vice versa). For example, the values of parameters change from ($\alpha 1$, $\beta 1$) to ($\alpha 2$, $\beta 2$) as the tab-chart moves from background (stage) to jump (stage) in standard sample. This classification method is not practical for the researchers and not automatic because it needs the human adjustment of parameters beforehand. However, this method has two purposes. Firstly, this method is a way to develop another automatic classification method. Secondly, this method is used as a tool to analysis the data reduction and fitted error of HOLT model.

The second algorithm of detecting the point change by HOLT model called classification rules. This algorithm is an automatic method because it uses a set of rules to divide the tab-chart into different stages. The classification rules are developed from classification method in the following way. After the tab-chart is well fitted by the HOLT model, the graphs of trend estimate versus smooth estimate are plotted for all tab-charts. The rules are drawn by comparing the trend-smooth graphs with the tab-charts. The background stage of the tab-chart has small trend and smooth values in the trend-

smooth graphs. The jump stage of the tab-chart has larger trend and smooth values. The plateau stage of the tab has larger smooth values but small trend values. The drop stage of the tab-chart has negative and large trend values. The performance of the classification rules are verified and tested by applying to the tab-charts of standard samples. Some guidelines of evaluating the performance are made to minimize the personal and biased judgement. Different person possibly has different judgement and view on some borderline cases. The classification rules have the successful rate ranging from 45% to 80% in various elements. However, after excluding the tab-charts of having close background and plateau, the successful rate of the classification rates will be at least 65%. This project provides the good method and platform of identifying the point change. One promising way of improving the performance is to refine and modify the rules. Although the rules seem to play more important role in the performance than the parameters (i.e. α and β), an experiment should be carried out to investigate any effect of the parameters on the performance.

For both classification method and rules, there is a problem of misclassification. This problem is that the classification and rules have the terrible performance in some tab-charts. In other words, there are a lot of observations in these tab-charts being wrongly classified. However, all these tab-charts have the property that the level of background and plateau are very close. Thus, this is possibly the cause of misclassification but the gentle signal change (i.e. jump or drop having gentle slope) is another possible cause. Anyway, this problem gives us a hint to improve the performance of both method and rules. For example, another set of rules is needed to tackle these tab-charts. Besides, classification method is better than classification rules because rules are hardly to replace the human visual judgement. However, classification rules are automatic and more practical than classification method.

Data reduction is another characteristic of the HOLT model. The model is capable of removing the noise or variation from tab-chart. The researchers can gain the useful information by observing the fitted curve (values) of the model. Apart from the graphs, quantitative analysis is also included in this thesis to help us to understand the data reduction in another angle. The square-root of the sample size formula is used to calculate the standard error over the background and plateau. This formula has the assumption of independent data. Although many tab-charts have the autocorrelation because the ARIMA model can be fitted to them, this violation does not cause the problem of using the formula. The formula is not used to estimate the statistical summaries of the underlying process. It is used to measure the degree of variation and fluctuation in the background and plateau. Over the plateau, the standard error of HOLT model (i.e. mean of fitted curve or values) is smaller than that of tab-chart (i.e. mean of the actual observation). This supports that the HOLT can reduce the variation over the plateau. However, the situation is reversed over the background. The standard error of HOLT model over background is greater than that of tab-chart over background. It implies that the HOLT model has difficult in background and plateau of small signal (i.e. trace amount of element). A constant trend (i.e. more smooth fitted curve) should be used on the background by choosing the appropriate

parameters. In other words, the background (or plateau of small signal) should use different α and β parameters.

After fitting the time series model to the raw data, the analysis of the fitted error should also be provided. From the analysis result, the fitted error of HOLT model increases as the level of plateau increases. The fitted error of the model decreases as the mass of the element increases. It is because the amount of heavy elements is smaller than the amount of light elements in the sample. Therefore, the only factor of affecting the fitted error is the level of plateau. This result implies that the standard deviation of fitting error of element's concentration can be minimized if the background noise can be controlled or minimized. Thus, the control of background noise can help to estimate the trace of the element in the sample more accurate but it is not capable of delivering only significant improvement to the major element in the sample.

For a comprehensive analysis, ARIMA model should also be fitted to the tab-chart. Since there are a lot of tab-charts to be fitted, a policy is devised to speed up the ARIMA model fitting. The procedures of fitting the ARIMA have the following three steps. The first step is to check the tab-chart is stationary. If it is not stationary, differencing will be carried out on the background of the tab-chart. The second step is fit AR(1), AR(2), MA(1), MA(2) and ARMA(1,1) to the background, the best model is chosen by the lowest MSE and significant parameters. If the five models cannot be fitted to the background, it may be random walk or has another ARIMA model. The third step is to use the above two steps to fit the ARIMA to the plateau of the tab-chart. After the model fitting, the result is obtained at the following. Over the background, no differencing is needed and most of the tab-charts are ARMA(1,1), random walk or inconclusive. Over the plateau, most of the plateaus having upward or downward trend are ARIMA(0,1,1), whereas most of the plateaus having horizontal trend are AR(1) or ARMA(1,1). The other plateaus are random walk or inconclusive. The above result also indicates that most of the tab-charts have the autocorrelation problem.

In the method (model) selection, there are several reasons of not choosing ARIMA model to tackle the researchers' problem. Firstly, the equation and structure of ARIMA is too complicated to be modified to identify the point change. Secondly, although simple exponential smoothing model is a special case of ARIMA model (equivalent to ARIMA(0,1,1) model without constant term), this model is not accepted in the selection. The reason is already mentioned in the previous paragraph. Thirdly, the researchers are looking for an automatic and fast method to extract the information from the tab-chart, ARIMA model seems to be not a practical method for automated use. However the ARIMA can be used to detect the autocorrelation existing in the tab-chart before applying the HOLT model. The pre-process on the tab-chart could be done if the autocorrelation is serious and the precise estimate is required to make the judgement on the tab-chart or sample.

There are some inadequate places in this project and more work is needed on these places in further studies. Firstly, the classification rules are relatively approximated and should be refined. More advanced mathematical techniques could be employed to devise better classification rules. Secondly, the other models should also be explored and investigated such that they have better performance in change point detection and the researchers can gain more information from the tab-chart via these models, for example, locally weighted regression. Thirdly, this project does not have enough work on studying the parameters of the HOLT model. One way is to investigate the impact of parameters on the performance of classification rules because only one set of parameters is used in this analysis. Another way is to investigate the relationship between parameters and data reduction because the data reduction does not work on the background or plateau of small signal. Lastly, some studies should be done to tackle the tab-chart having the autocorrelation and influence of autocorrelation on the statistical estimate on the tab-chart by HOLT model.

Moreover, there are several directions worthwhile to be considered in the long-term goal. Firstly, the classification rules can be extended and applied to the tab-chart of multiple significant signals (i.e. jumps and drops) and multiple flat parts (i.e. backgrounds and plateaus). Secondly, when studying the standard samples in this project, the tab-chart of each element in a sample is investigated independently and separately. However, the mineral in the real sample is a chemical compound of elements. Therefore, the change of composition in real samples will involve the investigation of more than one tab-chart. Thirdly, the knowledge (i.e. model and method) of this thesis is not limited only on the composition investigation of minerals or rock samples. It can be generalized and applied to other areas (i.e. charts in other problem). Lastly, the HOLT model and other methods of change point analysis should be compared, especially their performances. Since the data reduction of HOLT model is to trace the trend of the tab-chart and remove the variation or noise, the HOLT model may be incorporate into other methods to get better performance.

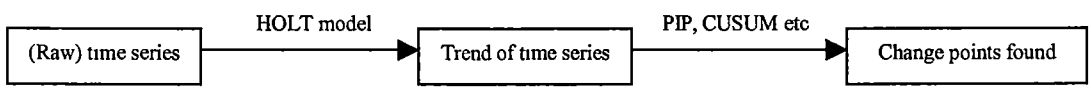


Fig 10.1 : Incorporate data reduction technique of HOLT model into other methods

Table Of Contents

Chapter 1 Introduction And Overview

(1.1) Preamble

Laser ablation – inductively coupled plasma – mass spectrometry (LA-ICP-MS) is a relatively new advance in instrumental multi-element low-level analysis that has been largely developed for geological research purposes, but is finding wider application to the *in situ* inorganic analysis of many solid materials [22], [23].

The LA-ICP-MS is a method to analyse the composition of minerals and the tab-charts will be produced from the ICP-MS instruments, which show the distribution and amount of elements in the mineral (see Section 1.8). In this research, the *data reduction technique* is applied to filter out the noise of the tab-charts and trace the trend of the tab-charts.

(1.2) Structure Of This Thesis

The first four chapters (i.e. Chapter 1 to Chapter 4) are informative and instructive. These chapters give the basic background information to the readers and help them to understand the work of this project in later four chapters (i.e. Chapter 5 to Chapter 8). The content in the first four chapters and Appendix A mainly come mainly from an unpublished paper [9], and additional materials (including paperwork and Excel spreadsheet) provided by CODES and the School of Earth Sciences, University of Tasmania and websites [16], [25] and [26]. These materials have been synthesised and rewritten for thesis presentation.

(1.3) Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry

Laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) is a new technique for measuring the inorganic composition of the minerals and other types of solid samples [16]. The CODES instrument has some special features [16], [25] which include (a) small amount of samples is sufficient and required for fully composition analysis; (b) the instrument can tackle very small-sized and any solid samples; (c) up to sixty elements can be analysed at one time; (d) the trace elements in the sample can be quantified and measured; (e) in some cases, heavier isotopes can be measured with good precision.

(1.4) Laser Ablation Process (LA)

Laser ablation is a process of using a laser beam to detach material from a solid sample. The ablated material is transported in an Ar gas stream to the torch of an ICP-MS for compositional analysis [9], [16], [25], [26]. Details of this process will be discussed later (see Section 1.6).

(1.5) Application Of LA-ICP-MS

LA-ICP-MS has wider application in both academic and industrial areas [16], [25]. They include (a) the evidence collection of crime and forensics; (b) elemental analysis and quantification of mineral and rock samples; (c) the investigation of environment pollution; (d) the exploration of the trace elements in polymeric materials; (e) process and quality control; (f) reliability and failure analysis.

In CODES (ARC Centre of Excellence in Ore Deposits) and the School of Earth Sciences at the University of Tasmania, the staff and other researchers have been routinely analysing sulphide and, to a lesser extent, non-sulphide minerals for minor and trace elements by LA-ICP-MS for several years. The technique allows for the quantitative measurement of up to 60 elements *in situ* with a spatial resolution of a few tens of microns, or less. Further details of the instrumentation are discussed in Chapter 2.

(1.6) Two Types Of ICP-MS: LA Vs Solution

Before discussing the specifics of the CODES LA-ICP-MS instrument it is worth comparing LA- sampling with conventional (solution-based) techniques for presenting material to the ICPMS.

In LA analysis the material to be analyzed comprises very small particles with a range of sizes that are carried in a stream of Ar gas to the plasma of an ICPMS. In the plasma these particles are vaporized and ionized (see Section 2.3.1). Then the ionized species or ions are sampled and pass through a quadrupole mass discriminator (see Section 2.3.2) and on to a detector for counting. The detector measures the number of electrons / hits from the selected ions (see Section 2.3.3). Lastly, these numerical data are captured as a csv file for subsequent processing (see Section 4.1).

For solution-based ('conventional') ICP-MS analysis the material being analysed is digested (dissolved) and presented to the plasma in a dilute, weakly acid, solution.

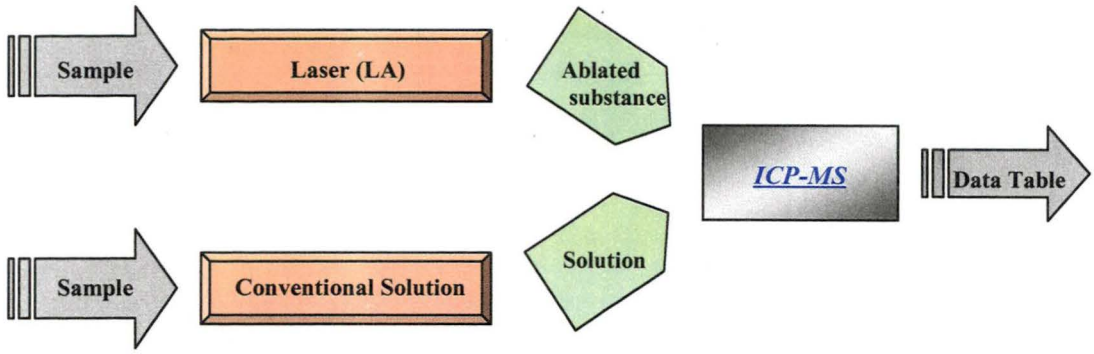


Fig 1.6 (1) : LA-ICP-MS And Conventional Solution-ICP-MS. The input is sample collected from field site, whereas the output is data table

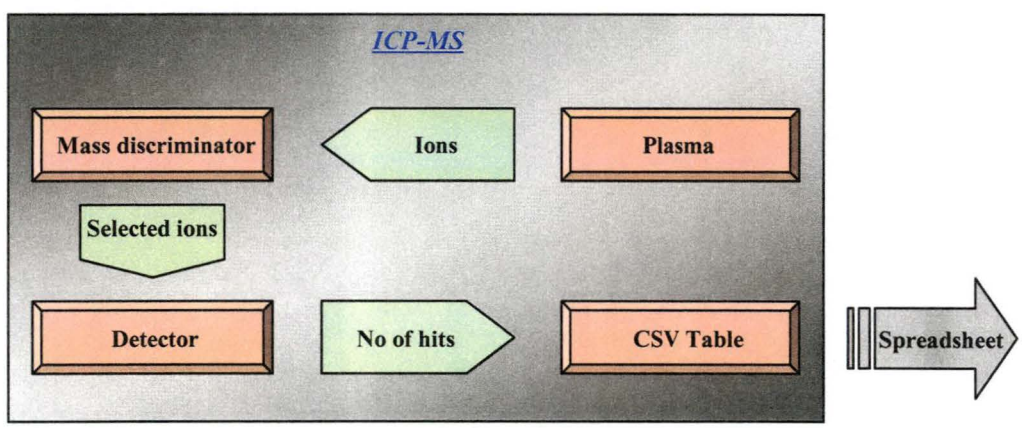


Fig 1.6 (2) : The Internal Structure Of ICP-MS. The output is spreadsheet

(1.7) Advantages Of LA-ICP-MS Over Conventional ICP-MS

LA-ICP-MS has some advantages compared with conventional solution-ICP-MS [16], [25]. Primarily, the excellent spatial resolution (see section 2.2) allows sampling and analysis of complex microscopic intergrowths common to many geological materials. Sampling can be done in real time, and sample preparation is minimal (usually a polished piece of rock is used). Lastly, contamination is usually not a problem with LA techniques.

By contrast, solution techniques require much larger amounts of material (rock or mineral powder). The powders need to be obtained by micro-drilling, or sub-sampling a larger crushed rock sample. Both procedures may lead to sample contamination, either from adjacent mineral grains in the case of drilling, or from crushing equipment with larger samples. Furthermore, some types of

geological materials are notoriously difficult to dissolve in mineral acids, thus incomplete digestion can also be a major problem with solution-based techniques.

(1.8) Data From The ICP-MS: The Tab-Chart

A tab-chart (e.g., Fig 1.8) is a time-series graphical representation of all measurements made by the ICP-MS instrument during the course of a single laser 'burn'. The x-axis represents the starting time of each cycle (see section 4.1), whereas the y-axis represents the counts per second or signal intensity (i.e. counts for a particular ion as measured by the detector of the ICP-MS). Initially, counts are low ("background") until the laser is switched on. Laser switch-on is reflected in the tab-chart by an abrupt increase in total counts for the elements of interest. The curves reveal an element's change as the instrument is applying the laser on the sample (see Section 4.2.1). Tab-charts obtained for 'unknown'-samples are then compared to tab-charts for 'known' materials (referred to as 'standards') in order to quantify individual trace elements in the 'unknown' sample being analysed (see Section 4.2.2, 4.2.3 and 4.2.4).

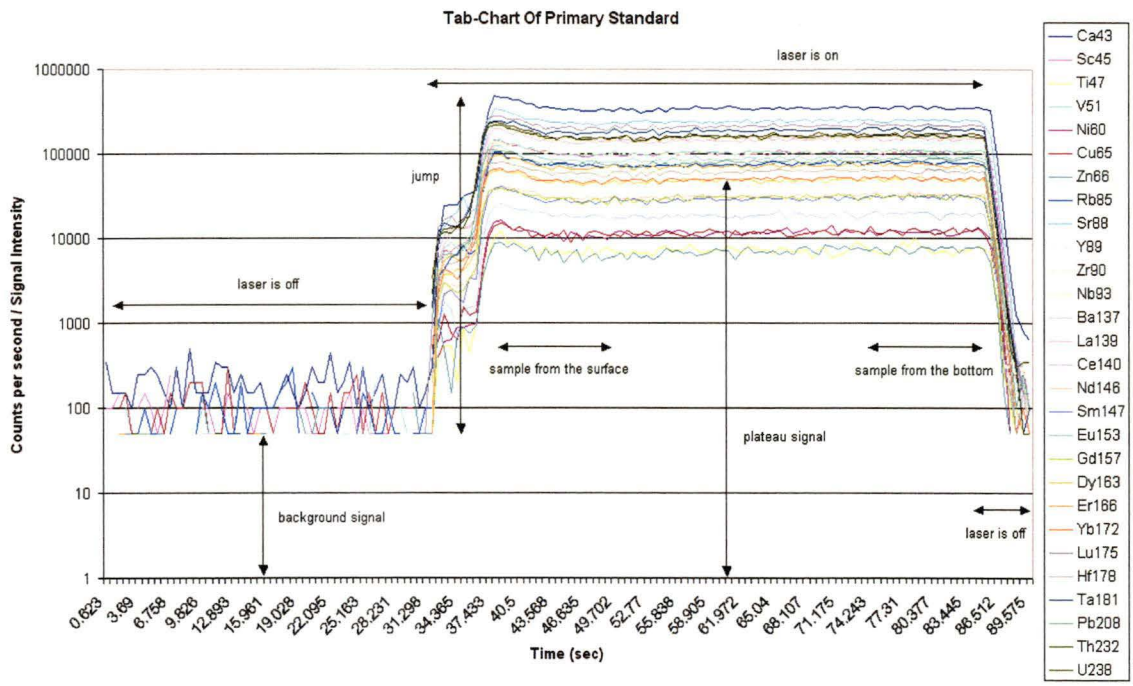


Fig 1.8 : *Tab-chart Of A Primary Standard*

(1.9) Standard Samples Studied In This Project

This thesis deals with LA-ICP-MS data obtained from two 'standard' glass materials. The first of these is a synthetic glass 'NIST 612' purchased from the US National Institute of Standards and Technology. This glass has been doped with sixty-one trace elements to a nominal concentration of 50 mg/kg for each trace element. The second is a glass was produced in the School of Earth Sciences by fusing a certified natural rock (basalt) powder (BCR2) sourced from the US Geological Survey. Over the last couple of decades both NIST 612 and BCR2 have been extensively analysed in many laboratories worldwide using many techniques. Hence, their composition is very precisely and accurately known for a large number of elements. CODES provided the raw data from sixty-six LA-ICP-MS spot analyses of these glasses (six of NIST 612 and sixty of BCR2).

For the purpose of quantification of LA-ICP-MS analyses, unknown samples are usually compared to a analyses of known materials such as NIST 612 and BCR2. These 'primary standards' are analysed at regular intervals along with the unknowns (typically, two spot analyses of a primary standard for every ten to fifteen unknowns). As well as quantification of unknowns, the regular analysis of primary standards allows instrumental drift to be accounted for during daily ICP operation. Sometimes more than one 'known' is run with a group of unknowns, and the second known ('secondary standard') becomes an additional check on analytical precision and accuracy. For more description and discussion on standard samples, see chapter 3.

(1.10) The Aims And Goals Of This Project

(1.10.1) The Main Aim And Goal

To date researchers identify and classify individual tab-charts by their observation, knowledge and past experience. This can be referred to as the 'visual' or 'intuitive' method. However there are grey areas that they are not easy to identify and classify by simply observation. Therefore, it would be useful to have some mathematical and statistical methods to aid with these judgements.

The main goal of the research described here is to find out systematic methods or rules to classify tab-charts into different portions. Ultimately, this may aid geological researchers by allowing tab-chart classification to be performed more quickly and automatically. The detailed work to this goal will be presented in chapters 5 and 6.

(1.10.2) The Other Aims And Goals

Apart from tab-chart classification, the researchers want to retrieve useful information from the tab-chart quickly and automatically. Chapters 5 and 6 will provide the knowledge (including equations and examples) how to get the information (i.e. statistical summaries) from the tab-chart. Chapter 7 includes the quantitative and statistical analysis to validate the method and model and evaluate their performance. This chapter also includes some work investigating factors affecting the magnitude of errors. Chapter 8 investigates the autocorrelation structure of the tab-charts and fits and criticize ARIMA models for the tab-chart segments.

(1.10.3) The Summary Of The Work In This Project

The work reported here concentrates on the data from the two glasses discussed in section 1.9, with NIST 612 referred to as a 'primary standard' and BCR2 as a 'secondary standard'. Tab-charts for 'unknown' samples are invariably more complex than those of glasses, and their statistical analysis is beyond the scope of the present study. The work of chapter 5 to chapter 6 is to find out the systematic methods or rules and give their statistical summaries (see sections 1.10.1 and 1.10.2). There are several works shown in chapter 7. Firstly, the noise of the tab-chart and our method will be estimated. Secondly, the error of our method will be estimated. Thirdly, the current method used by the researchers and our method developed in this project will be compared. In chapter 8, the ARIMA model will be applied to the tab-charts and some discussions are made about ARIMA model and our method. The summary of chapter 5 to chapter 8 is displayed at the following graph.

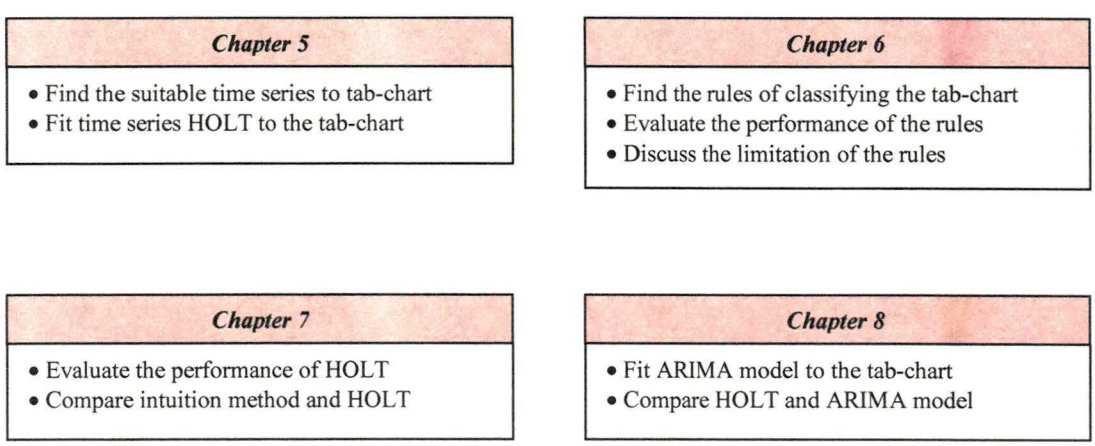


Fig 1.10 : *Summary Of Chapter 5 to Chapter 8*

(1.11) Literatures (Current Methods And Our New Method)

(1.11.1) Literatures Of Laser Ablation Process

The materials and information about laser ablation process can be referenced to [9], [22], [23], [25], [26] and [27]. The work of this project is based on the paper [9] and the raw data is provided by CODES and the School of Earth Sciences, University of Tasmania.

(1.11.2) Literatures Of Change Point Analysis

Change point analysis (detection) or *time series segmentation* is a problem in the data mining and this problem has two directions : the first direction is to detect the mean shifts in the time series and the second direction is to detect the variance change in the time series. For simplicity, the problem is how to divide (cut) a time series into segments. Each segment represents a mean shift in the time series. Change point analysis is familiar with *statistical process control* (SPC), which is used to monitor the process and detects any occurrence of *out of control* and abnormal pattern. SPC is well-known in quality and process control. The main difference between SPC and change point analysis is that the former focus on the detection of change point one at a time but the latter detects all change points at one time.

P. Hubert (2000) shows that the change points (mean shifts) in the time series can be modelled as a tree. The change points are found by choosing the segmentation in the least squares sense. The solution space is huge and *branch and bound* is impossible to tackle for moderate sized problem. So this paper uses the *Scheffe test* and some constraints to find the solution. Fang Li (2006) tackles the change points (mean shifts) in multivariate process by tree-based supervised learner. Chen (2008) uses the evolutionary method to solve the problem of change points (mean shifts). In his journal, he uses k-means and Euclidean distance to calculate the fitness of chromosome.

Taylor (2000) uses *cumulative sums of difference* (CUSUM) and *bootstrapping* to detect the change points (mean shift and variance change) in the time series. The idea is that a change in direction of the CUSUM chart indicates a change point in time series. However it is not easy to identify the change points by observing the CUSUM chart only. Thus, bootstrapping produces the bootstrap CUSUM charts to help to find out the change points. This method can be applied to both mean shift and variance change problem. Carslaw (2006) is a paper to apply the technique CUSUM and bootstrapping to urban air pollution concentration time series.

Fu (2006) uses *perceptually important points* (PIP), a method is applied to technical patterns in finance and stock market, to identify the change points (mean shift) in the time series. The time series is re-arranged into their importance by PIP identification process. The most important point is

the first change point and the second important point is the second change point in the tine series and so on. Then a binary tree is constructed and the change points are retrieved from the tree according to their importance. Kumar (2001) incorporate fuzzy set theory into the change point analysis. This article states that the change point (mean shift) often occur over a time interval and should not be treated as a sudden change at a particular time.

The linear Holt exponential smoothing model is chosen in this paper and applied to the tab-chart. Since the tab-chart is actually a time series, the useful signal and information can be reflected and extracted by fitting the model to the tab-chart. In this thesis, this model is modified and can be used to divide the tab-chart into different segment. Comparing with the other methods of change point analysis, the modified model has two advantages. Firstly it is easily implemented in the Excel spreadsheet. Secondly, apart from the function of change point detection (i.e. time series segmentation), the modified model can also provides the statistical information (summaries) and estimate error of each segment. The model is also capable of removing the variation and noise from the tab-chart. Thus, the researchers can gain the useful and comprehensive information in the fast and efficient way.

The modified model is applied to the standard sample in this thesis because it is artificial and made in the laboratory. The situation becomes more complicated in the real samples. However, the study of standard samples can build the good foundation and platform for studying real samples. It offers the prospect of building up knowledge about tab-chart and understanding the advantages and disadvantages of the modified model.

Chapter 2 Structure Of LA-ICP-MS Instrument

(2.1) CODES Laser Analytical Facility

The CODES Centre of Excellence in Ore Deposits at the University of Tasmania operates a laboratory equipped with laser sampling (Nu Wave UP-213) and ICP-MS (Agilent Technologies 4500 plus) instruments, which are used to analyse the composition of rock and mineral samples. This chapter briefly describes these instruments and their operation.

(2.2) Nu Wave UP-213 Laser Ablation System

This part of the system (Fig. 2.2 (1) and Fig 2.2 (2)) allows careful targeting with high spatial resolution of small areas of (polished) rock and mineral samples. The sample chamber accommodates a single 2.5 cm diameter polished mount that can contain unknown samples or standards. A video camera and microscope allow real-time imaging of the mount and laser ablation run. A computer-controlled, motorized X-Y-Z stage allows precise movement (± few microns) of the sample to allow selection of laser targets. The laser software (MEOLaser-213) controls the laser parameters including spot size, laser energy, pulse rate and firing of the laser and ablation takes place under a stream of He gas that carries the ablated material away and towards the ICP-MS.

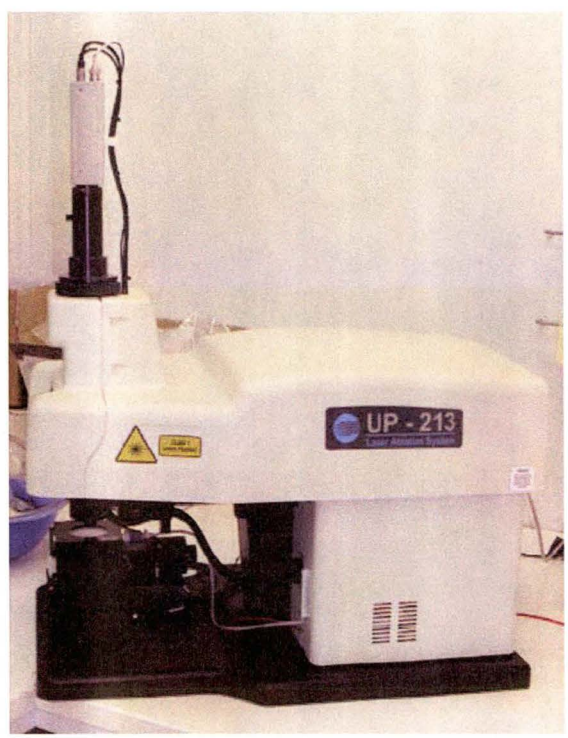


Fig 2.2 (1) : *UP-213 Laser Ablation System*

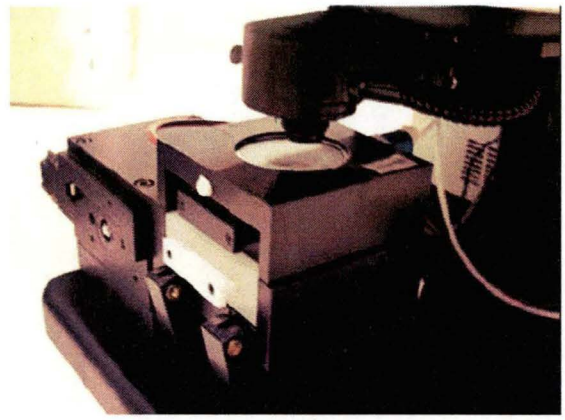
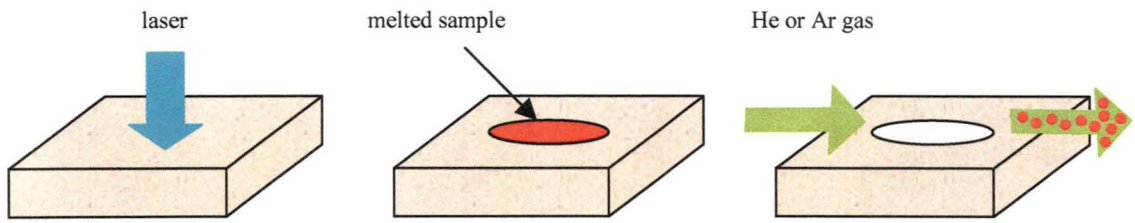


Fig 2.2 (2) : *Sample Chamber And Motorised Stage*

(2.2.1) ***The Function Of Laser and Gas***

The laser is pulsed and interacts with the surface of the rock sample. In a complex process, partly dependent on the nature of the sample, tiny amounts of material are ablated and a hole is created. He and Ar gas carry the ablated material to the plasma of the machine ICP-MS. There are two types of laser operation on the rock sample : spot (vertical) analysis and line analysis.



Step 1 : Laser hit the sample surface *Step 2 : Sample is melted* *Step 3 : He gas entrain the sample*

Fig 2.2 (3) : *The Job Of Laser And Gas In Sample Chamber*

(2.2.2) ***Spot (Vertical) Analysis***

In spot mode the laser can produce a hole ranging from 8 to 110μm in diameter. The laser is pulsed (5 to 10 Hz) and continues to hit on the same spot during the process. However, as the hole deepens (typically 10 to 30 μm, it becomes more difficult for the carrier gas to entrain material from the bottom of the hole. This will cause the decrease in the signal intensity. Spot analysis allows for easy removal of surface contamination by pre-ablating the spot for a few cycles of the laser.

(2.2.3) Line Analysis

In line analysis, sample is moved beneath the laser as analysis proceeds. This produces a shallow linear sampling track across the sample, instead of a single hole. Dealing with surface contamination in line analysis can be problematic.

(2.3) Agilent HP4500plus Quadrupole ICP-MS

The ICP-MS instrument is responsible for analysing the composition of ablated material. It is made up of three main parts : plasma, mass discriminator and detector. In detail, a silica glass torch and induction coil produces an Ar plasma when the instrument is running; twin metal cones with 'pinholes', just beyond the plasma, allow ions to pass into the second (high vacuum) part of the instrument which is a quadrupole mass filter (Fig. 2.3 (3)); and ions selected by the quadrupole filter are then counted by an electron multiplier device that can operate in both pulse (for low count rates) and analog (high count rate) modes. Numerical data from the detector are tabulated by the ICP-MS controlling software as csv tables suitable for export to Excel or other software packages.

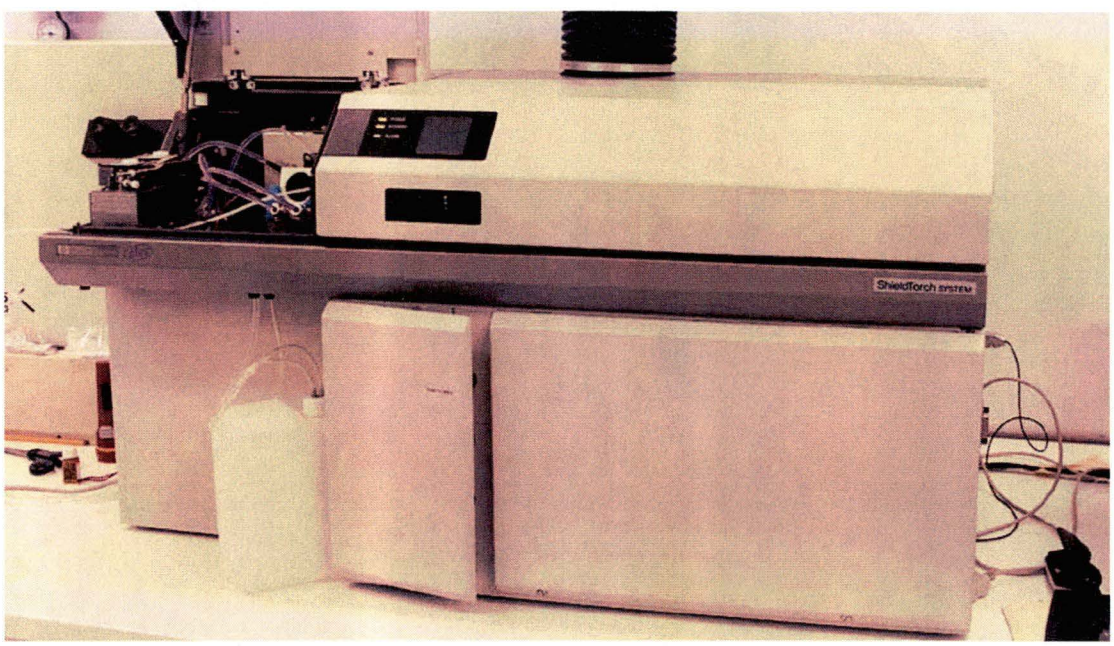


Fig 2.3 (1) : *Mass Spectrometer / ICP-MS*

(2.3.1) ICP Torch Body / Plasma

The ICP torch body three concentric silica glass tubes wrapped by an induction coil. The coil is used to transmit radio frequency energy to the Ar gas flowing through the three tubes (Fig 2.3 (2)). The plasma in the central (injector) tube of the torch may reach temperatures of over 6500°K. Such a plasma is very efficient at breaking chemical bonds and creating simple (singley-charged) ionised species.

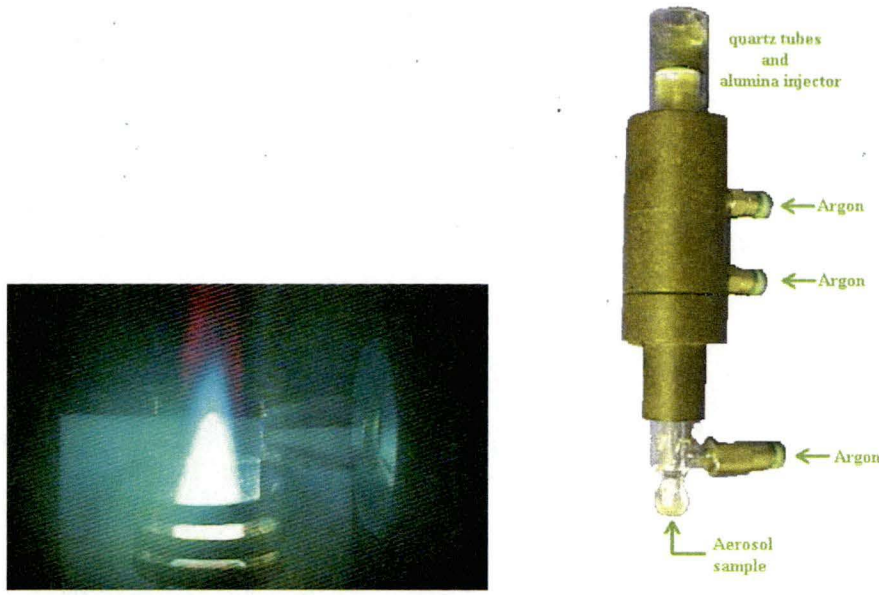


Fig 2.3 (2) : *Plasma (Source From http://ewr.cee.vt.edu/environmental/teach/smprimer/icpms/icpms.htm)*

(2.3.2) ***Mass Discriminator Or Quadrupole Mass Filter***

Mass discriminator or Quadrupole mass filter is a vacuum chamber. This vacuum environment prevents the collisions between ionised species and air molecules. This chamber contains four magnetic rods ('quadrupole'), which is used to select the specific ionised species for detection. By changing the intensity of magnetic field, only selected species will reach the detector on other side of the chamber, whereas other species are driven away from the chamber based on their mass to charge ratio.

Fig 2.3 (3) : *Mass Filter (Source From http://ewr.cee.vt.edu/environmental/teach/smprimer/icpms/icpms.htm)*

(2.3.3) ***Detector***

The detector is an electron multiplier (Fig. 2.3 (4) and (5)) and the last major component in the machine ICP-MS. Positive ions reaching the detector are attracted to it and release electrons when they impact the first part of the detector, these in turn impact the next part of the detector to produce more electrons. This 'multiplier effect' helps account for the high sensitivity of ICP-MS analysis.

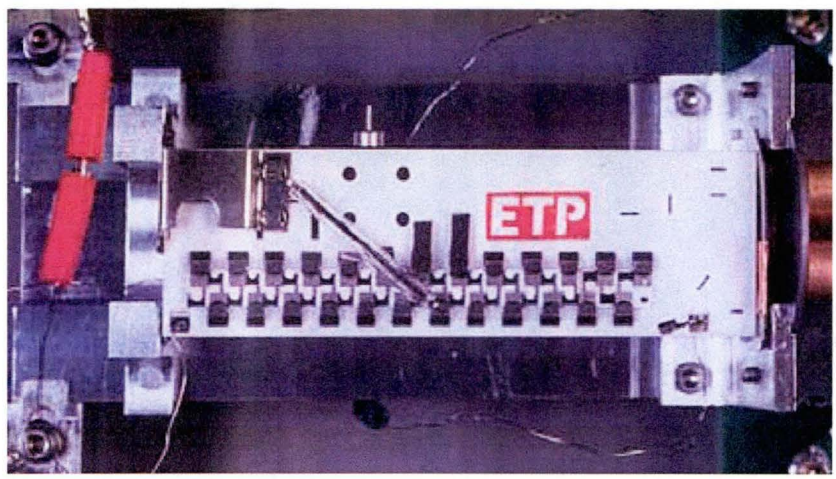


Fig 2.3 (4) : *Detector (Source From http://minerals.cr.usgs.gov/icpms/intro.html)*

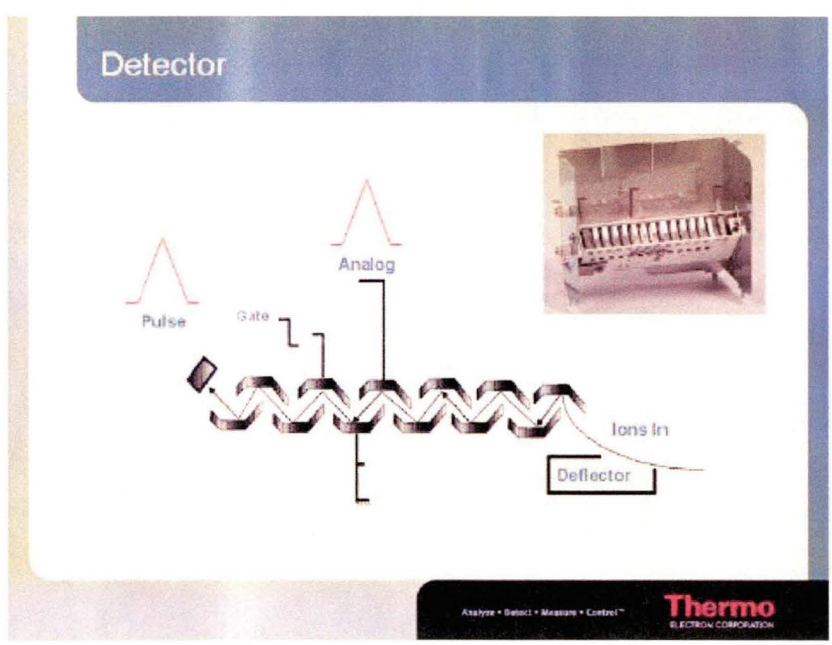


Fig 2.3 (5) : *Detector (Source From http://www.uea.ac.uk/env/technical/lab/icp-ms2.shtml)*

(2.3.4) ***Sample Composition Investigation***

In combination, the quadrupole mass filter and the detector rapidly scan the isotope masses being measured (~ 0.5 seconds to scan and count 30 different masses). Counts are accumulated for each mass and recorded. In a typical LA analysis ablation and counting would continue for about 100 seconds, generating more than 150 scans across the mass spectrum (there is time 'lost' while the detector is inactive between scans). Data for each selected mass is captured as counts per second (cps) for each scan. These machine generated data tables form the basis for all subsequent calculations and quantification of LA-ICP-MS data.

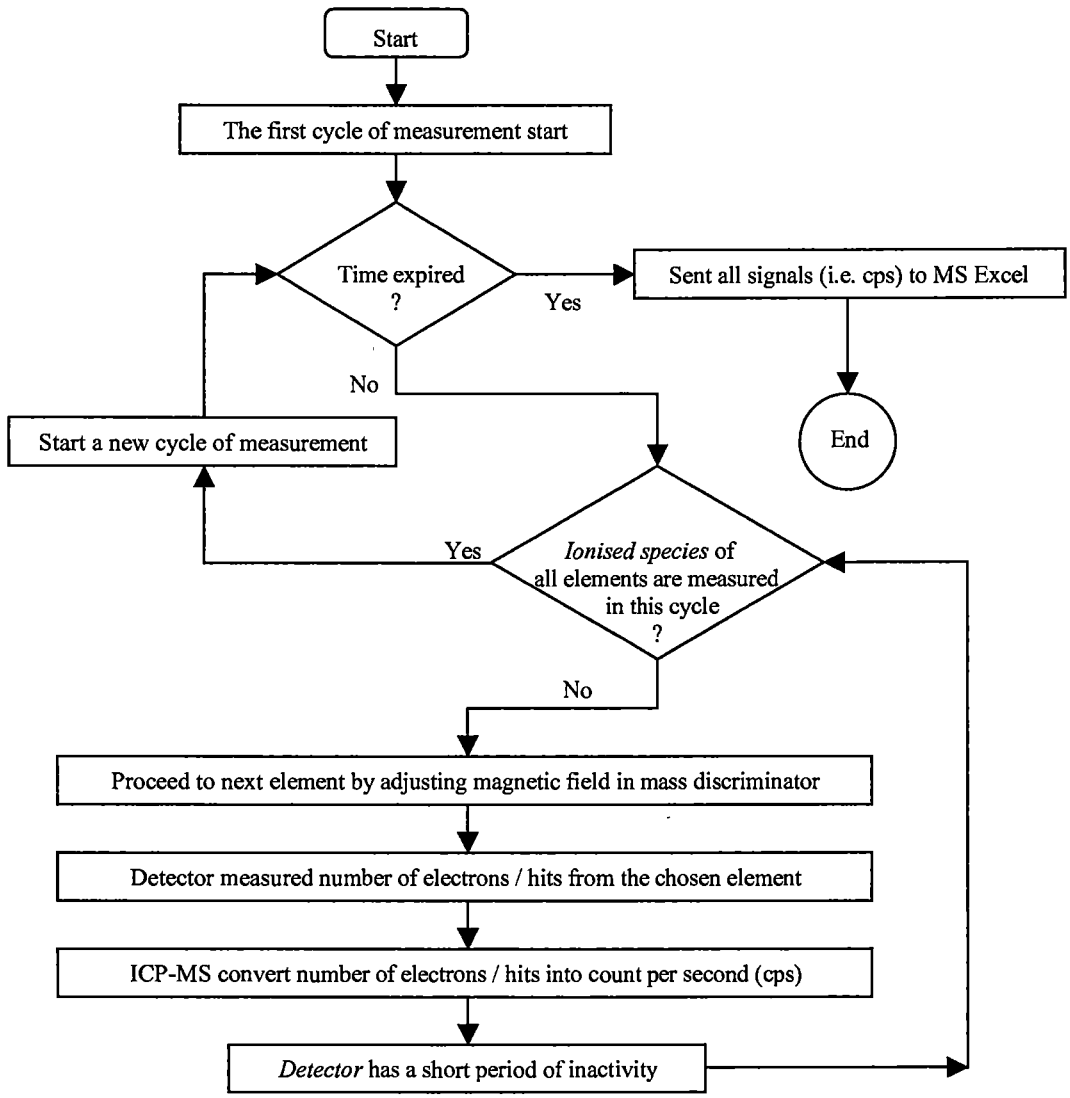


Fig 2.3 (6) : *Detector Job In ICP-MS Machine*

Chapter 3 Materials Used For This Study

(3.1) Standards

Quantification of ICP-MS analyses is made by comparing measurements made on unknown materials to measurements made on well (compositionally) characterised natural or artificial materials under similar analytical conditions. These 'known' materials are referred to as standards. For solution-based ICP-MS the standards are solutions made to contain precisely known concentrations of the elements of interest. For LA-ICP-MS standards are either synthetic glasses 'doped' with a variety of trace elements, or natural materials fused to a homogeneous glass. In the latter case, the composition of the glass has been agreed by multiple-interlab and multiple-instrument procedures. Glasses make good LA-ICP-MS standards because, at the scale of laser sampling, they are homogenous and trace elements are assumed to be uniformly distributed.

For the work undertaken herein LA-ICP-MS data from two glass standards (NIST610 and BCR-2) have been used. NIST610 is a synthetic silicate glass doped with a nominal 50 ppm of 61 elements prepared by the US National Institute of Standards and Technology (www.nist.gov) and Fig 3.1 (1). BCR-2 is a basalt powder prepared by the US Geological Survey (minerals.cr.usgs.gov/geo_chem_stand/basaltbcr2) – Fig. 3.1 (2). A small aliquot of BCR-2 was fused in a platinum crucible and quenched to an homogenous glass at the laboratory at the School of Earth Sciences, University of Tasmania.

(3.2) Primary Vs Secondary Standards

For quantification purposes (see Chapter 4), the standard used to calculate elemental concentrations in LA-ICP-MS analysis is referred to as the 'primary' standard. Sometimes, in order to check analytical precision and accuracy during routine LA-ICP-MS runs, another glass standard is analysed and treated as an 'unknown'. This second glass standard is referred to as a 'secondary' standard.

Primary standards are analysed before and after the analysis of unknown samples (pre- and post-analysis). Commonly, two measurements of the standard glass bracket each set of 10-12 unknowns. By bracketing measurements it is possible to correct for machine drift (see Chapter 4).

The CODES and the School of Earth Sciences provides sixty-six standards to this project. Six standards are primary standards (samples 1, 2, 33, 34, 65 and 66) and the other samples are secondary standards. There are twenty-eight elements in each standard. Therefore there are total 1848 (= 66 samples × 28 elements in each sample) tab-charts for the studies in this project.

Fig 3.2 (1) : *Composition of Standard Glass - NIST612. Corrected to six significant figures*

Element	Count per second	Proportion (%)	Element	Count per second	Proportion (%)
Li7	41.54	0.000077	Ru101	0	0
Be9	37.73	0.000070	Rh103	0	0
B11	34.73	0.000065	Pd105	0	0
Na	98740.5	0.184184	Ag107	21.92	0.000041
Na2O	13.31	0.000025	Cd111	28.32	0.000053
Mg24	77.44	0.000144	In115	42.93	0.000080
MgO	0.012843	0.000000	Sn118	37.96	0.000071
Al	10531.6	0.019645	Sb121	38.34	0.000072
Al2O3	1.99	0.000004	Te125	0	0
Si	339876	0.633984	Cs133	41.64	0.000078
SiO2	72.72	0.000136	Ba137	37.74	0.000070
P39	55.16	0.000103	La139	35.77	0.000067
K39	66.26	0.000124	Ce140	38.35	0.000072
K2O	0.007982	0	Pr	37.16	0.000069
Ca43	84691	0.157978	Nd146	35.24	0.000066
CaO	11.85	0.000022	Sm147	36.72	0.000068
Sc45	41.05	0.000077	Eu151	34.44	0.000064
Ti49	48.11	0.000090	Gd157	36.95	0.000069
TiO2	0.008022	0.000000	Tb	35.92	0.000067
V51	39.22	0.000073	Dy163	35.97	0.000067
Cr53	39.88	0.000074	Ho	37.87	0.000071
Mn55	38.43	0.000072	Er166	37.43	0.000070
Fe57	56.33	0.000105	Tm	37.55	0.000070
FeO	0.007247	0.000000	Yb173	39.95	0.000075
Co59	35.26	0.000066	Lu175	37.71	0.000070
Ni60	38.44	0.000072	Hf178	34.77	0.000065
Cu65	36.71	0.000068	Ta181	39.77	0.000074
Zn66	37.92	0.000071	W182	39.55	0.000074
Ga69	36.24	0.000068	Re185	8.12	0.000015
Ge72	34.64	0.000065	Ir193	0	0
As75	37.33	0.000070	Pt195	0	0
Se77	14.44	0.000027	Au197	5.09	0.000009
Rb85	31.63	0.000059	Tl205	15.07	0.000028
Sr88	76.15	0.000142	Pb208	38.96	0.000073
Y89	38.25	0.000071	Bi209	29.84	0.000056
Zr90	35.99	0.000067	Th232	37.23	0.000069
Nb93	38.06	0.000071	U238	37.15	0.000069
Mo95	38.3	0.000071			

Fig 3.2 (2) : *Composition of Standard Glass BCR-2. Corrected to six significant figures*

Element	Count per second	Proportion (%)	Element	Count per second	Proportion (%)
Li7	9.20	0.000017	Ru101	0	0
Be9	2.30	0.000004	Rh103	0	0
B11	0	0	Pd105	0	0
Na	23603.3	0.042895	Ag107	0	0
Na2O	3.18167	0.000006	Cd111	0	0
Mg24	22249.9	0.040435	In115	0	0
MgO	3.69	0.000007	Sn118	2.5	0.000005
Al	71948.4	0.130754	Sb121	0.28	0.000001
Al2O3	13.595	0.000025	Te125	0	0
Si	251441	0.456950	Cs133	1.17	0.000002
SiO2	53.7983	0.000098	Ba137	682	0.001239
P39	1608.22	0.002923	La139	25	0.000045
K39	14721.3	0.026753	Ce140	53.7	0.000098
K2O	1.77333	0.000003	Pr	6.78	0.000012
Ca43	50957.6	0.092607	Nd146	28.6	0.000052
CaO	7.13	0.000013	Sm147	6.6	0.000012
Sc45	32.6	0.000059	Eu151	1.95	0.000004
Ti49	13434.4	0.024415	Gd157	6.68	0.000012
TiO2	2.24	0.000004	Tb	1.05	0.000002
V51	413	0.000751	Dy163	6.3	0.000011
Cr53	14.7	0.000027	Ho	1.28	0.000002
Mn55	1548.91	0.002815	Er166	3.61	0.000007
Fe57	96249.6	0.174915	Tm	0	0
FeO	12.3822	0.000023	Yb173	3.38	0.000006
Co59	39	0.000071	Lu175	0.5	0.000001
Ni60	12.1	0.000022	Hf178	4.9	0.000009
Cu65	23	0.000042	Ta181	0.81	0.000001
Zn66	125	0.000227	W182	0	0
Ga69	22	0.000040	Re185	0	0
Ge72	0	0	Ir193	0	0
As75	0	0	Pt195	0	0
Se77	0	0	Au197	0	0
Rb85	46.8	0.000085	Tl205	0.29	0.000001
Sr88	333	0.000605	Pb208	10	0.000018
Y89	33.4	0.000061	Bi209	0	0
Zr90	183	0.000333	Th232	6.05	0.000011
Nb93	12	0.000022	U238	1.7	0.000003
Mo95	244	0.000443			

(3.3) Internal Standards

During the course of each individual (standards and unknowns) measurement (usually 60-100 seconds) systematic variations in signal intensity can occur for several reasons unrelated to 'real' changes in the composition of the material being analysed. For quantitative ICP-MS analysis, an 'internal standard' is used to account for these fluctuations. An internal standard can be any constituent of the material being analysed for which accurate independent compositional information is available. For solution ICP-MS analysis of solutions, internal standards are usually one or more elements added to the solution in known amounts. For LA-ICPMS the internal standard element is usually a major element of the mineral being analysed previously measured by electron microprobe or by assuming mineral stoichiometry. For example, elements Fe and S are the internal standard elements used in pyrite analysis, whereas elements Pb is the internal standard elements for galena.

Chapter 4 Tab-Charts And Calculations

(4.1) Raw Data

The ICP-MS detector measures the raw data of signal / counts per second (cps) for one isotope at a time, with the mass spectrometer rapidly cycling through a pre-selected list of isotopes (elements). This means a single laser run of 60 to 80 seconds produces numerous short counts for each isotope. Counting takes place while the detector 'dwells' on each selected isotope (typically 20-30 milliseconds, depending on the number of isotopes selected). There is a 'settling' time of a few milliseconds between each mass jump. The ICP-MS software preserves the counts and some other instrument parameters as csv data tables which are imported into Microsoft Excel for further manipulation. There are three uses for these data sheets : plotting 'tab-charts', calculating element concentrations and calculating analytical error [9].

Figure 4.1 (1) and (2) has examples of these csv tables presented as an Excel spreadsheet. Due to the limited space, only small portion of two spreadsheets, containing the raw data of secondary sample 25 and 12, are shown in the following graphs. Actually, there are total 29 columns and 149 rows in each spreadsheet because 146 cycles are done in the primary and secondary sample. The column represents the readings (i.e. cps) of an element in all cycles, whereas the row represents the readings (i.e. cps) of all elements in each measurement cycle. These readings are then used and plotted as the tab-chart (see Section 4.2.1).

By looking into the first spreadsheet at below, the row 4 represents the first measurement cycle. This cycle starts at time 0.624. The ICP-MS detector measures the cps of the elements in the sequence Ca43, Sc45, Ti47, V51,....., U238. After this isotope finishes counting, the detector stops for a while (i.e. settling time). A second the second cycle (i.e. row 5) starts at time 1.237 and repeats the same measurement process. The last cycle occurs at time 89.576 and the measurements are recorded at row 149. For example, the ICP-MS detector measures that the cps of V51 at measurement cycle 140 is "52311.19".

	A	B	C	D	E
1	C:\DATA\050309\MA09C25.D				
2	Intensity Vs Time	CPS			
3	Time [Sec]	Ca43	Sc45	Ti47	V51
4	0.624	450	0	0	50
5	1.237	250	0	0	0
6	1.85	450	0	0	0
143	85.899	11953.19	5250.61	130580.8	52311.19
144	86.512	11953.19	4800.51	135057.3	55619.17
145	87.126	10802.61	4050.37	132995	55769.54
146	87.739	10802.61	5550.69	126758.8	55869.79
147	88.353	11753.09	4250.4	128016	52762.25
148	88.966	10902.65	4300.41	128820.6	57323.47
149	89.576	11202.8	4350.42	126256	51208.64

File name and sample 25

For each cycle, the machine measures elements from left to right one by one

The third cycle starts at time 1.85

This cycle starts at time 85.899

In this cycle, the machine measures that the cps of V51 is "52311.19"

Figure 4.1 (1) : *Raw Data Of Sample 25 Compiled In Microsoft Excel*

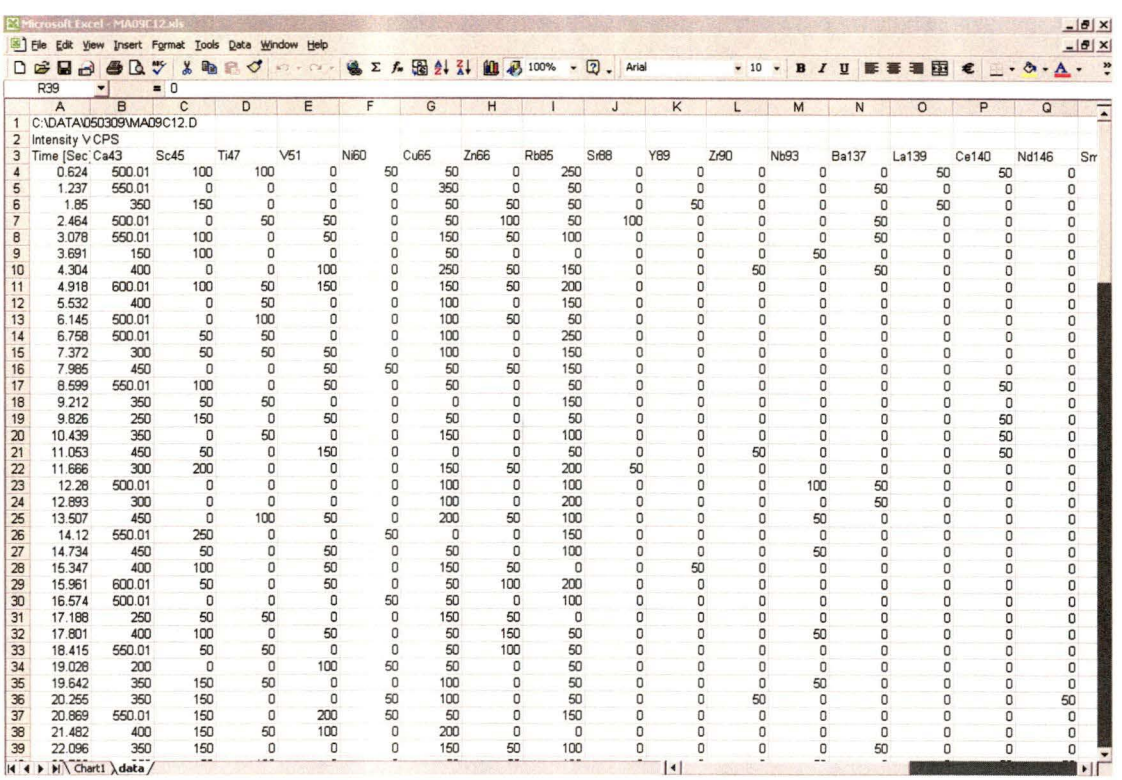


Figure 4.1 (2) : *Raw Data Of Sample 12 Compiled In Microsoft Excel*

(4.2) Tab-Chart Description

(4.2.1) Structure Of Tab-Chart

A 'tab-chart' is a graphical, time-series representation of all measurements made by the ICP-MS instrument during the course of an individual spot or line analysis. Figure 4.2 (1) show a tab-chart of primary standard. A single analysis usually takes 90 to 120 seconds, with the first 30 seconds used to measure the background from the Ar and He carrier gas and the plasma. When the laser is switched on ablated material is rapidly carried to the plasma. This causes a sudden jump to much higher count rates for the elements being measured. Quite quickly, this signal stabilises to a flat pattern for the remainder of the ablation, only dropping when the laser is switched off. These two parts of the tab-chart are labelled jump and plateau on Figure 4.2 (1).

The plateau portion of tab-chart is used for concentration calculations. The plateau is subdivided into one or more portions called integration interval. Each integration interval has consistent trend or slope, assumed to represent homogenous phases or zones in the sample.

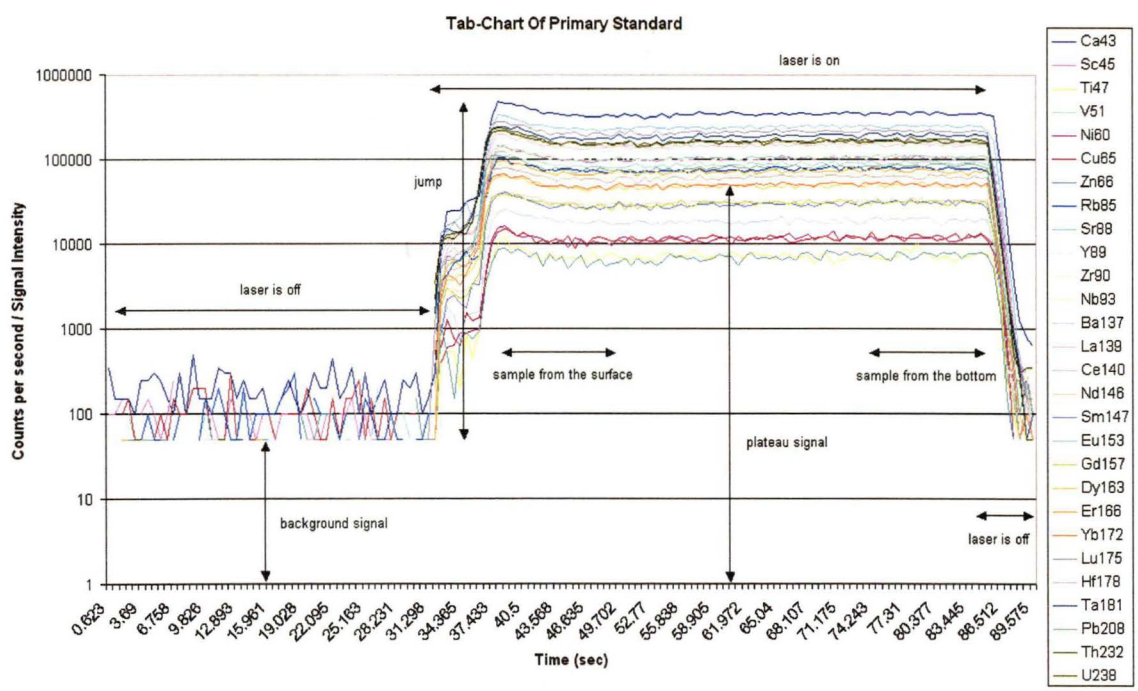


Figure 4.2 (1) : *Tab-Chart Description*

(4.2.2) ***Investigation Of Element Composition In Tab-chart***

In natural heterogeneous materials the different zones can be identified by the trend and change of signal in the tab-chart. If there is a change in the composition of major elements, it indicates that another mineral has been reached in the sample. Figure 4.2 (2) shows an example of a single analysis accidentally ablating two minerals in one 'burn'. The spot ablation commenced in pyrite (high Fe but low Cu), then penetrated into chalcopyrite (increase in the signal from Cu).

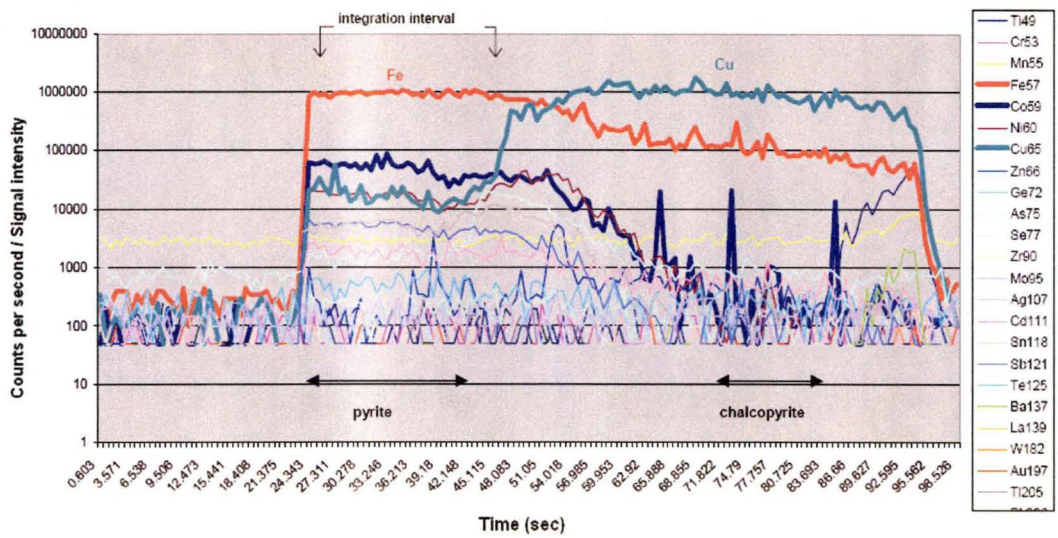


Figure 4.2 (2) : *Tab-Chart Of A Sample Having Different Minerals (Source From [9])*

By looking into tab-chart pattern, researchers can obtain a qualitative indication of the proportion and distribution of various elements in unknown samples. Figure 4.2 (3) is a sample of pyrite with generally flat and parallel patterns for the elements Pb208 and Bi209. This indicates fairly evenly distributed micro-inclusions of galena (PbS) in the sample, or an even distribution of Pb and Bi substituted in the pyrite lattice. By contrast, the very 'spiky' Pb and Bi in pyrite shown on Figure 4.2 (4) indicate large discrete galena inclusions are present in the sample.

Geological workers currently identify the spikes and signals by visual inspection and categorisation of the tab-charts. They need to decide these spikes and signals are caused by the mineral inclusion, systematic changes in a mineral's composition or simply random noise or other unknown reasons. Therefore, one of this project's aims is to find systematic and automatic methods to classify these spikes and signals. One approach is to use the statistical methods / tests to test their significances. If these spikes or signals are significant when challenged by these tests, we can conclude that they are caused by the 'real' features of the material being analysed. Otherwise they are caused by random noise or other unknown reasons and must be included in concentration calculations.

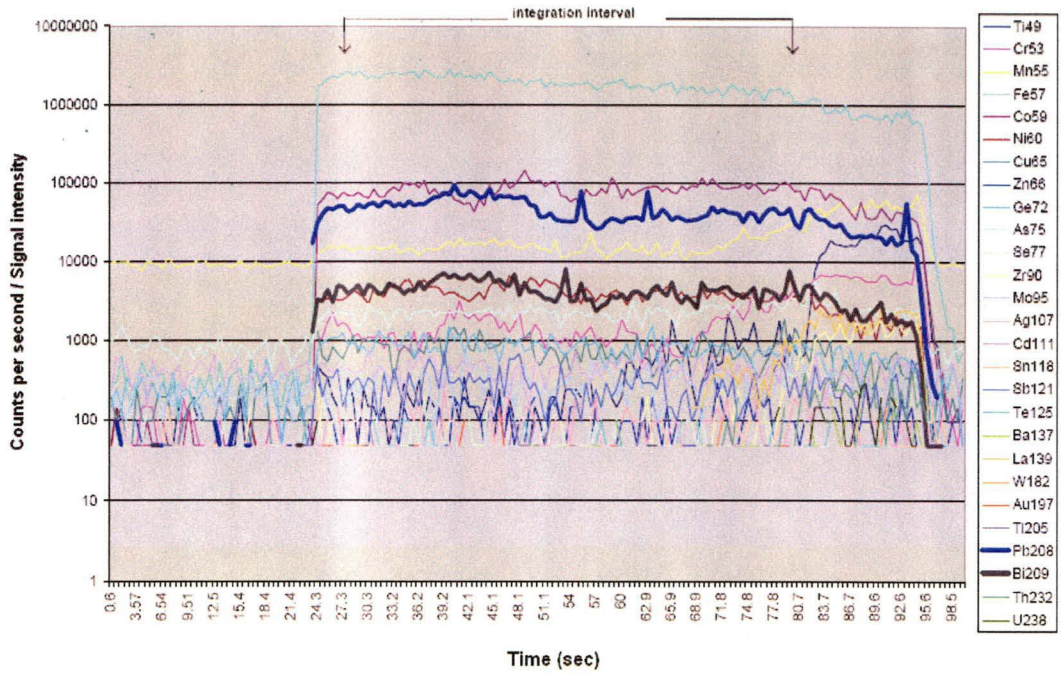


Figure 4.2 (3) : *Tab-Chart Of Pyrite Sample Having Evenly Distributed Micro-Inclusions Of Galena (Source From [9])*

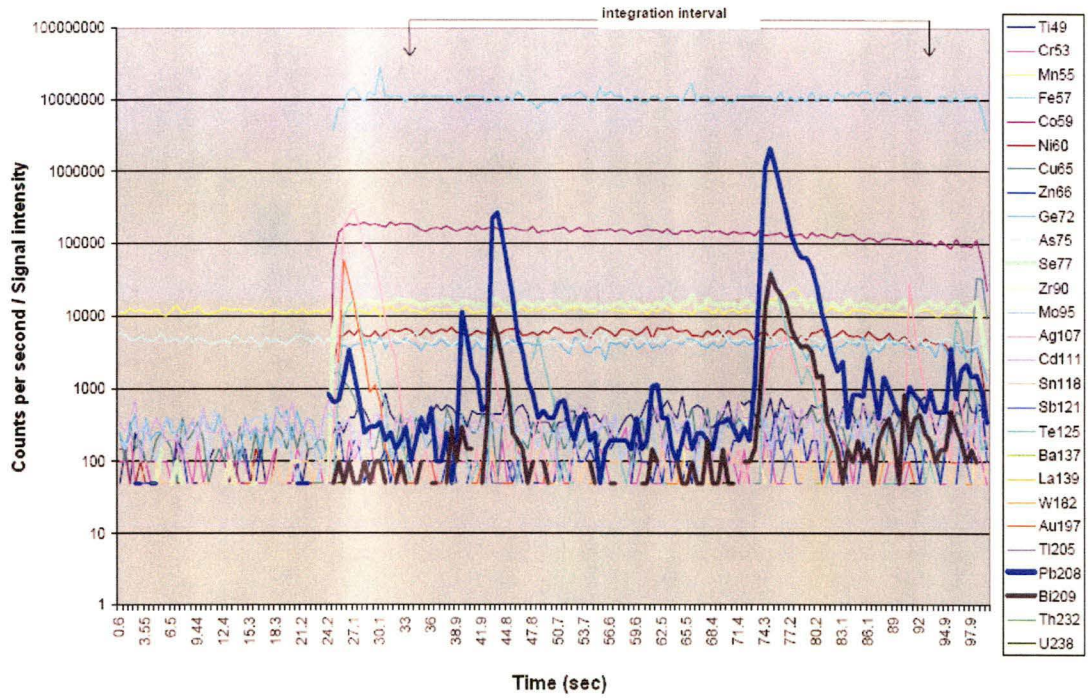


Figure 4.2 (4) : *Tab-Chart Of Pyrite Sample Having Discrete Galena Inclusions (Source From [9])*

(4.2.3) ***Influence Of Spot-Size On Tab-Chart***

The laser can be focused to produce holes ranging from 110 microns to as small as 8 microns in diameter. With smaller the spot-sizes the narrower the hole produced will cause a down-hole decrease in the signal intensity as it is harder for the gas flow to extract ablated material from the bottom of the hole. The following figure shows that there is a continuous drop in signal when the hole is small.

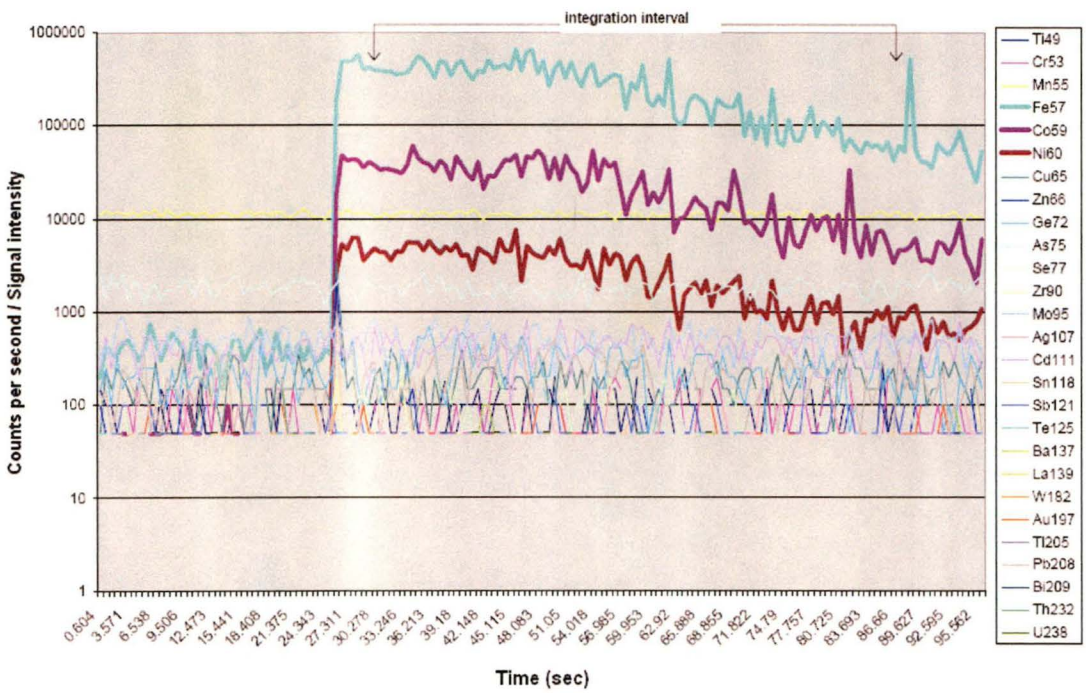


Figure 4.2 (5) : *Tab-Chart Of A Sample With Small Hole (Source From [9])*

(4.2.4) ***Influence Of Contamination On Tab-Chart***

If the sample has not been pre-ablated (i.e. to remove surface contamination) in spot analysis, an initial spike will occur when the laser is fired (i.e. at the start of plateau in the tab-chart). This contamination should not be included in the integration interval. The following figure shows the shape of a contaminated sample without pre-ablation.

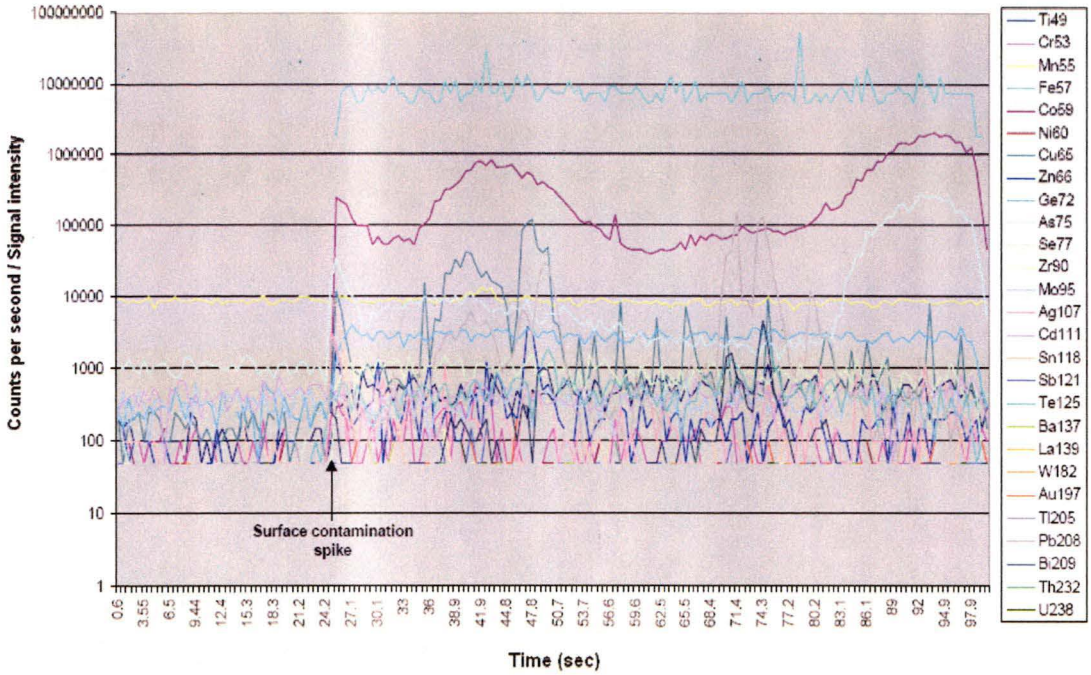


Figure 4.2 (6) : *Tab-Chart Of A Contaminated Sample (Source From [9])*

(4.3) Steps In Calculating Element Concentrations

(4.3.1) The Selection Of Integration Interval

The first step is to identify and select the integration intervals for background and plateau. For the background (plateau), the researchers use their visual judgement, experience and knowledge to decide the starting point and ending point of the background (plateau) in the tab-chart. There is a remark that the terms *count per second, signal (intensity)* and *observation value* are the same meaning in this thesis.

Once they are selected, the observation values of background and plateau are averaged over their integration intervals. In other words, the researchers calculate the average value of the background (plateau) by averaging the observation values over the integration interval for background (plateau). Then the average amount of element ($\overline{cps}_{\Delta SGBG}$) is obtained by subtracting the average counts per second of background ($\overline{cps}_{BG}$) from the average counts per second of plateau ($\overline{cps}_{SG}$).

(4.3.2) The Machine Drift

The second step is to calculate the drift using the paired primary standard analyses run at within the group of 'unknowns'. These are used to calculate the quantities X_{pre} and X_{post} since the primary standard samples are analysed before and after unknown samples analysis. The equations of X_{pre} and X_{post} are given in the following. The definitions of the notations in the equations can be found in the appendix A (see appendix A2.2). It is noticed that X_{pre} and X_{post} are constant because the values come from the primary standard samples (see section 3.1).

$$X_{Pre} = \frac{1}{m}\sum_{t=1}^{m}\left(\frac{\left(\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)_{STD(t)}\right)^{E}}{\left(\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)_{STD(t)}\right)^{IE}}\right) \quad \text{and} \quad X_{Post} = \frac{1}{m}\sum_{t=m+1}^{2m}\left(\frac{\left(\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)_{STD(t)}\right)^{E}}{\left(\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)_{STD(t)}\right)^{IE}}\right)$$

Thus the drift of the machine D can be estimated by the proportional method because the researchers of the laboratory assume that the drift changes linearly proportional with time. In my opinion, the assumption of drift changing linearly is made because the machine is regularly calibrated and corrected. If the machine is not calibrated, the error will be incurred in the machine and it will give the inaccurate measurement to the researchers. The equation of the drift is given in the following. The definitions of the notations in the equations can be found in the appendix A (see appendix A2.2). The

term $(D)^{E}_{US(t)}$ is constant in the unknown sample t because X_{pre} and X_{post} are constant and the last term of the equation is a function of t.

$$(D)^{E}_{US(t)} = X_{Post} + (X_{Pre} - X_{Post}) \times \frac{\text{Total Unknown Sample Analysis} + m - \text{Unknown Sample t} - 0.5}{\text{Total Unknown Sample Analysis} + m}$$

(4.3.3) ***The Calculation Of Element Concentration***

The last step is to calculate the concentration of the element. Thus with the unknown samples their element concentrations can be calculated by comparing the average counts per second of plateau for standard and unknowns. The equation of calculating concentration of element is given in the following. The definitions of the notations in the equations can be found in the appendix A (see appendix A2.3).

$$(ppm)^{E}_{US(t)} = \frac{(\overline{cps}_{\Delta SGBG})^{E}_{US(t)}}{(D)^{E}_{US(t)} \times \left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)^{IE}_{US(t)}}$$

In unknown sample t, the denominator of the above equation is constant because the first term is $(D)^{E}_{US(t)}$ and the second term is the known value of internal standard element (see section 3.3). Therefore, the conclusion is that the concentration of the element (i.e. ppm) in a sample is linearly proportional to the amount of the element in a sample (i.e. $cps_{\Delta SGBG}$). In the form of equation, we have

$$ppm = \lambda \cdot cps_{\Delta SGBG} = \lambda(cps_{SG} - cps_{BG}) \quad \text{where } \lambda \text{ is a constant}$$

To avoid the confusion and mix-up of some terms about element, the definitions of these terms are provided in the following table.

Figure 4.3 : *The Definition Of Terms About The Element*

Term	*Notation*	*The Meaning Of The Term*
level	cps_{BG} (cps_{SG})	The signal intensity of the element in the background (plateau)
amount	$cps_{\Delta SGBG} = cps_{SG} - cps_{BG}$	The amount of element in the sample
concentration	$ppm = \lambda(cps_{SG} - cps_{BG})$	The concentration of element in the sample

(4.3.4) The Summary Of Calculating Element Concentration And Remarks

All the steps for calculating the concentration are summarized in the following flowchart. There are two remarks before finishing this section. Firstly, this section is only to explain the steps of calculating the concentration of the element. Derivation of the equations is not included because it is not main concern of this thesis. Secondly, some notations and equations in this section and appendix A are reconstructed from the results and data of the Excel worksheet, which is provided by CODES and the school of Earth Sciences, University of Tasmania. The worksheet and the unpublished technical paper [9] show the theory and concept and provide some equations only.

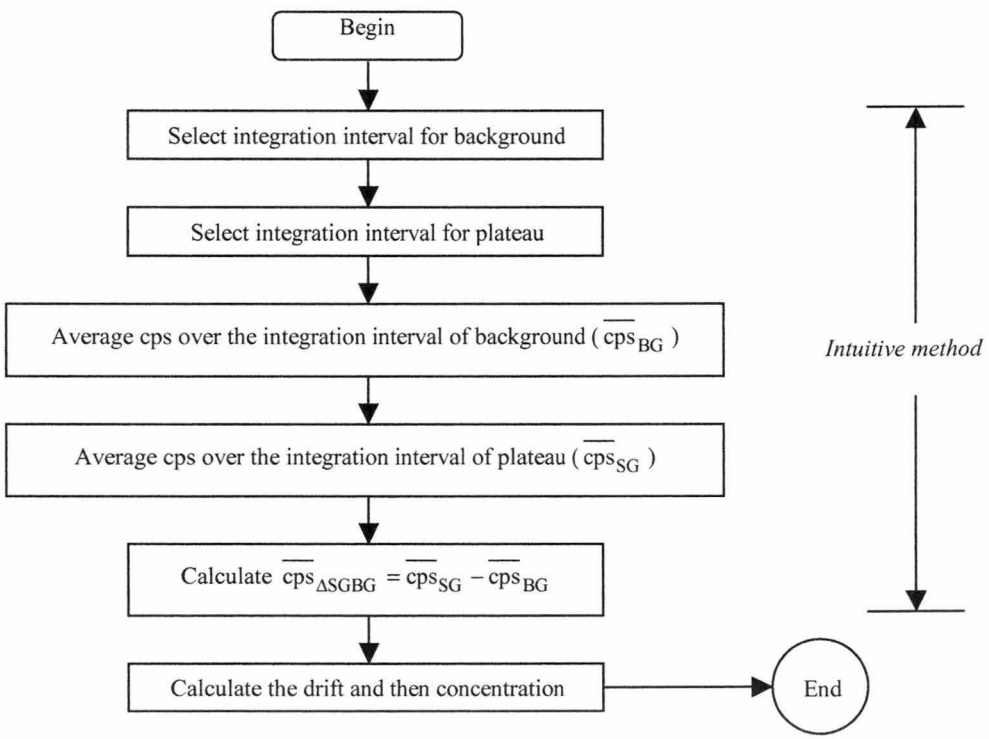


Figure 4.4 : *Flowchart Of Concentration Calculation*

(4.3.5) ***Intuitive Method***

There is no special term or name is given to the steps of the selection of integration interval and statistical calculations (summaries) over background and plateau in the above section (see section 4.3.1) in the unpublished technical paper [9]. For convenience, the name *Intuitive method* is given to these steps after discussion with my two supervisors. The statistical summaries of the Intuitive method are given in the following. One of the goals in this thesis is to explore an automatic and systematic method to replace the Intuitive method (see section 1.10.1 and Chapters 5 and 6). Two examples of the calculation of the statistical summaries are shown in the appendix (see Appendix F and G).

$$\begin{cases} \text{Mean (average level) of the background} : \bar{x}_{BG} = \sum_{k\in\Phi} \dfrac{x_k}{n_{BG}} \\ \text{Mean (average level) of the plateau} : \bar{x}_{SG} = \sum_{k\in\Omega} \dfrac{x_k}{n_{SG}} \\ \text{Standard deviation of the background} : \sigma_{BG} = \sqrt{\dfrac{\sum_{k\in\Phi}(x_k - \bar{x}_{BG})^2}{n_{BG}-1}} \\ \text{Standard deviation of the plateau} : \sigma_{SG} = \sqrt{\dfrac{\sum_{k\in\Omega}(x_k - \bar{x}_{SG})^2}{n_{SG}-1}} \end{cases}$$

where

(a) x_k = the count per second (i.e. signal intensity or observation value) of the tab-chart

(b) Φ = observations in the tab-chart are classified as background by intuitive method

(c) Ω = observations in the tab-chart are classified as plateau by intuitive method

(d) n_{BG} = total number of observations is classified as background by intuitive method

(e) n_{SG} = total number of observations is classified as plateau by intuitive method

Apart from the statistical summaries of Intuitive method, the researchers also need to know and calculate the statistical summaries of the element concentration. The statistical summaries (i.e. mean and standard deviation) of element concentrations are given by the following equations. When calculating the standard deviation of the element concentration, there is an assumption that raw data (i.e. observations) of background and plateau are independent.

$$\begin{cases} \text{Mean (average level) of the element concentration} : \lambda(\bar{x}_{SG} - \bar{x}_{BG}) \\ \text{Standard deviation of the element concentration} : \lambda\sqrt{(\sigma_{BG}^2 + \sigma_{SG}^2)} \end{cases}$$

(4.4) Detection Limit (DL) Calculation

The detection limit of element in unknown sample t is the lowest concentration that produces a statistical significant signal above the background. On the other words, this limit is the lowest concentration that we can distinguish from the background concentration. The detection limit equation is given at the appendix A.

(4.5) Total Analytical Error

The total analytical error of an element in unknown sample t (i.e. the total error of calculating the concentration) is given at the appendix A.

(4.6) Closing Remark

For the rest of this thesis the issue of quantification of unknown sample element concentrations will be ignored. From now the only data used in this thesis will be from the well characterised homogenous glasses and I will be primarily concerned with the statistical assessment of the shape of these simple tab charts.

Chapter 5 Modelling The Tab-Chart Of Secondary Standard Sample

(5.1) The Work Of This Chapter

This chapter will discuss the feasibility and suitability of different time series to the tab-chart. Some modifications are probably made before applying these time series to the tab-chart. After the suitable models are chosen, they are tried to fit to the tab-chart.

(5.2) Time Series Models To Tab-Chart

We want to develop a suitable and realistic time series model including meaningful parameters so that we can understand the measurement process, including its errors. However we are not using the time series to predict because we are not interested in predictions, but rather in parameter estimates. Optimality (or nearly) of the smoothing parameter choices should be driven by goodness of fit rather than by prediction.

The one-parameter and one-estimate (i.e. one equation) time-series (eg. simple exponential smoothing model, etc) is not suitable for the tab-chart of secondary standard sample because there are different trends in the different portions of the tab-chart. The equation of simple exponential smoothing model is at the following.

$$\text{Smooth} : L_t = \alpha Y_t + (1-\alpha)L_{t-1}$$

where

α = smoothing level parameter

$0 \le \alpha \le 1$

It seems that this one-parameter and one-estimate model is not adequate to explain the variation of the tab-chart. The tab-chart can be modelled as the following equation.

$$\text{Model} : Y_t = \mu_t + \gamma_t t + \varepsilon_t$$

where

μ_t = average mean / level of the tab-chart

γ_t = trend of the tab-chart

ε_t = fluctuation of the tab-chart

Although the three-parameters and two-estimates (two equations) time-series (eg. damped-trend linear exponential smoothing model, etc) can also be considered, it may involve three-dimensional plots and rather complicated in the beginning. The equations of damped-trend linear exponential smoothing are at the following.

$$\text{Smooth} : L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + \gamma T_{t-1})$$

$$\text{Trend} : T_t = \beta(L_t - L_{t-1}) + (1-\beta)\gamma T_{t-1}$$

where

α = smoothing level parameter

β = trend level parameter

γ = dampening parameter

$0 \le \alpha \le 1, \quad 0 \le \beta \le 1 \quad \text{and} \quad 0 \le \gamma \le 1$

Thus the two-parameters and two-estimates time-series (eg. linear Holt exponential smoothing model, etc) will be used for the first step. It has two estimates (i.e. trend and smooth) to reflect the variation of the tab-chart but has two parameters only. Besides, the seasonal time-series are not suitable and not considered since there is no obvious cyclic and seasonal pattern in the tab-chart. The ARIMA model will be discussed in Chapter 8. For more detail of various time series models, see [1], [2], [8], [10], [18] and [19].

(5.3) Linear Holt Exponential Smoothing Model

Due to the reason stated in the last section (see section 5.2), the linear Holt exponential smoothing model is good enough to fit the tab-chart of secondary standard sample even though this model allows the far-past to still play a part in estimates. This model has two parameters (i.e. α and β) : one is adjusted for the smoothing level and another one is adjusted for the trend. Then the estimating method (updating equations) is

$$\text{Smooth} : L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$\text{Trend} : T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

where

α = smoothing level parameter

β = trend parameter

$0 \le \alpha \le 1$ and $0 \le \beta \le 1$

Note that large α and β imply a large weighting on the current observations and a lesser weighting on the history. The k-step ahead predicted value is given by $\hat{Y}_t(k) = L_t + kT_t$. But our estimate of the current level is L_t (i.e. we use $\hat{Y}_t(0)$) and we use this as the fitted value. Since we desire to fit a model to the tab-chart and not to predict the future, $\hat{Y}_t(0)$ and $\hat{Y}_t(1)$ are used only in this thesis (see section 5.4.2).

(5.4) Modified Linear Holt Exponential Smoothing Model (HOLT Model)

(5.4.1) Equations Of Modified Linear Holt Exponential Smoothing Model

The linear holt exponential smoothing model cannot applied directly for tab-chart since it needs different parameters to explain each portion of the tab-chart : background, jump, plateau and drop. The reason is that background and plateau have small trends whereas jump and drop have large trends in this model. One parameter set should be used for background and plateau. Another parameter set should be used for jump and drop. Therefore, the model is modified such that it contains two pairs of parameters, say (α_1, β_1) and (α_2, β_2). The above parameter policy is only an initial trial and not perfect, the analysis result will reveal and reflect drawbacks and disadvantages of this policy.

In the modified linear exponential smoothing model, the equations are unchanged but an additional rule is used such that different parameters will change for each portion of tab-chart. The rule is listed as the followings : if absolute value of the trend, $|T_t|$, is less than a specified value, T, then (α_1,β_1) is chosen. Otherwise (α_2,β_2) is chosen.

$$(\alpha,\beta) = \begin{cases} (\alpha_1,\beta_1) & \text{if } |T_t| < T \\ (\alpha_2,\beta_2) & \text{otherwise} \end{cases}$$

The values (α_1,β_1), (α_2,β_2) and T will be adjusted manually such that this modified model seems fit for tab-chart. In the worksheet, the fitted curve (i.e. fitted value) of the HOLT model will change automatically when changing the values (α_1,β_1), (α_2,β_2) and T. By trial and error, the desired values are obtained when the fitted curve can trace the trend of the tab-chart. We hope that (α_1,β_1) can be chosen in the background and plateau portion and (α_2,β_2) can be chosen in the jump and drop portion according to the above rule. In other words, the parameter of this modified model is dynamic, whereas the parameter of the traditional model is stationary or unchanged when applying to the time series. For convenience, henceforth we use HOLT model as shorthand for this modified or dynamic-parameter model.

(5.4.2) ***Measurement Of HOLT Model's Performance***

We are actually interested in the current level and also the trend. We aim to use fitted value to characterize the stage or segment of the tab-chart (We are not interested in prediction). Two fitted values are used in this project. They are given at the following.

$$\text{(a) } \hat{Y}_t(0) = L_t \quad \text{and} \quad \text{(b) } \hat{Y}_{t-1}(1) = L_{t-1} + T_{t-1}$$

We will assess the goodness of fit of this method by the errors (i.e. $Y_t - \hat{Y}_t(0)$ or $Y_t - \hat{Y}_{t-1}(1)$), where Y_t is the observation. We do not expect this method to be optimal. We hope it compares satisfactorily to intuitive method.

(5.4.3) Fitting HOLT Model To Tab-Chart

First of all, the value T is adjusted manually to fit the tab-chart until a smooth curve is obtained and it can trace the trend of the tab-chart. Then the values (α_1,β_1), (α_2,β_2) will be adjusted for finding a better fitting curve for the tab-chart. The following diagram shows the tab-chart and HOLT model of secondary standard sample 3 in element Nb93, the parameters values and threshold value are $(\alpha_1,\beta_1) = (0.07,0.04)$, $(\alpha_2,\beta_2) = (0.80,0.20)$ and T = 40.00 respectively.

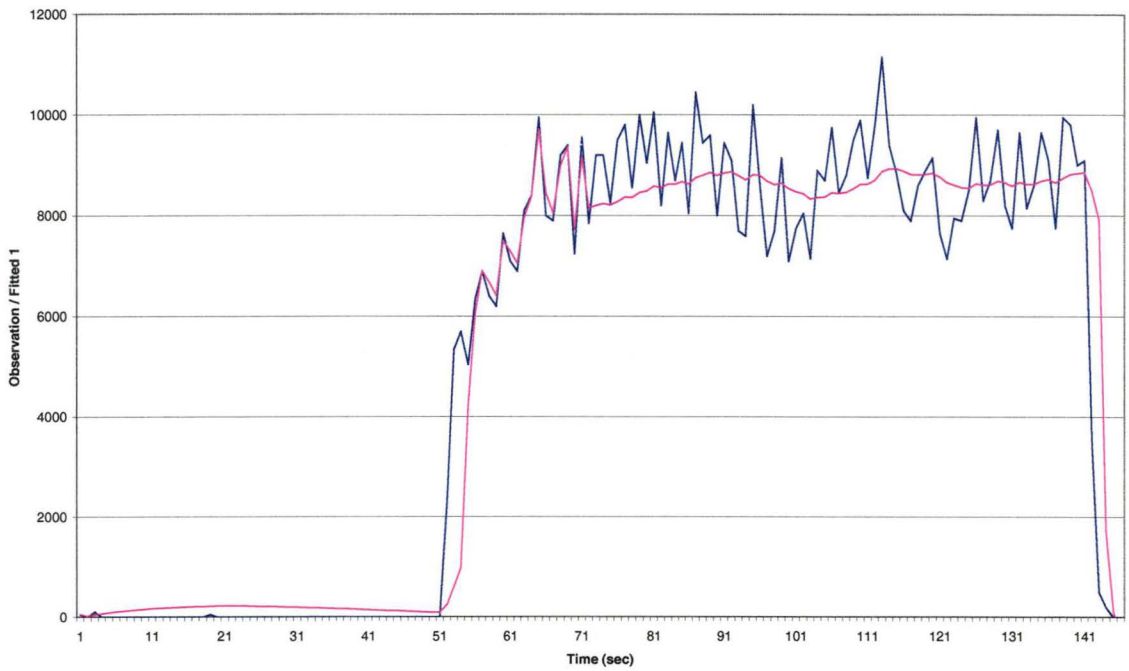


Fig 5.4 (1) : *Tab-chart And HOLT Model of Nb93, Sample 3. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 3 in element Nb93, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

After experience gained in processing all samples and elements, it is found that value T is more influential and important than values α and β in fitting time series to the tab-chart. If T is too small, the HOLT model is spiky in background and plateau although it is good fit to jump and drop. On the other hand, if T is too large, the HOLT model becomes smooth but it does not fit to the tab-chart well. As a result of this experience, (α_1,β_1) = (0.07, 0.04) and (α_2,β_2) = (0.80, 0.20) are chosen for all samples and elements.

The following three diagrams give a clear picture to support the above conclusions. The HOLT model is spiky at small T value (i.e. the first picture). As the value T increases, it is getting more and more smooth (i.e. the second picture). However, the HOLT model deviates from the tab-chart at large T value (i.e. the third picture). Notice that same T value has different result in different fitted curve (value) of the HOLT model. For example, T = 3.5, T = 9.0 and T = 50.0 will produce spiky, well-fitted and smooth curve in Cu65, sample 5 (see the following diagrams); whereas T = 2.0, T = 3.0 and T = 13.0 will produce spiky, well-fitted and smooth curve in Cu65, sample 25.

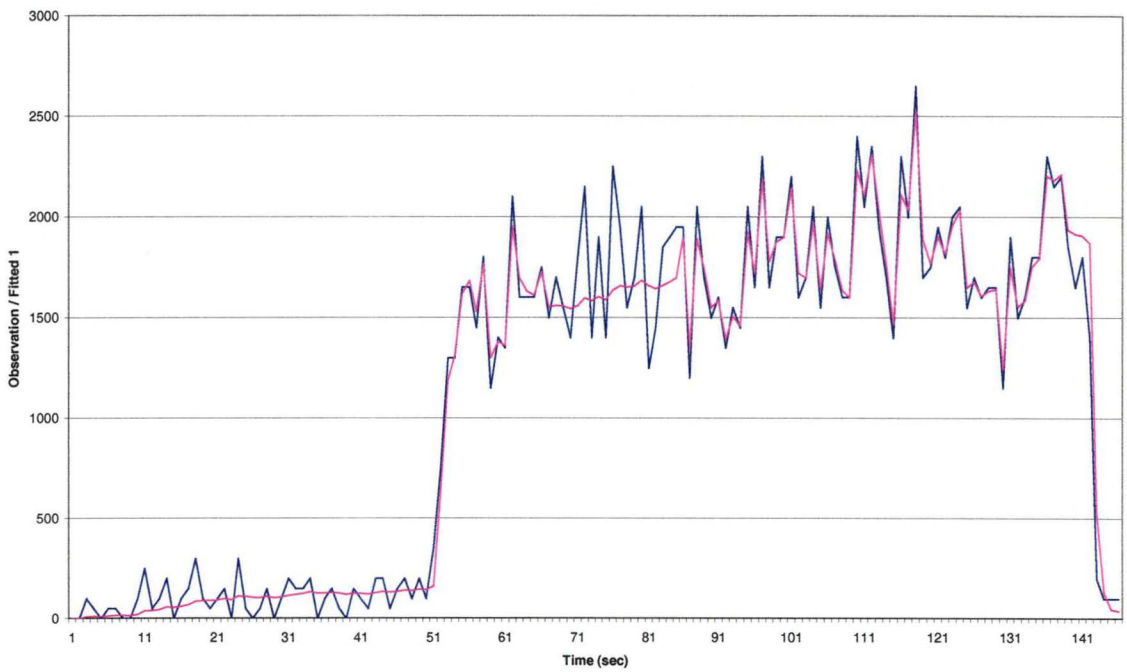


Fig 5.4 (2) : *HOLT Model Of Cu65, Sample 5. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 5 in element Cu65, whereas the smooth curve is the HOLT model with T = 3.5 (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

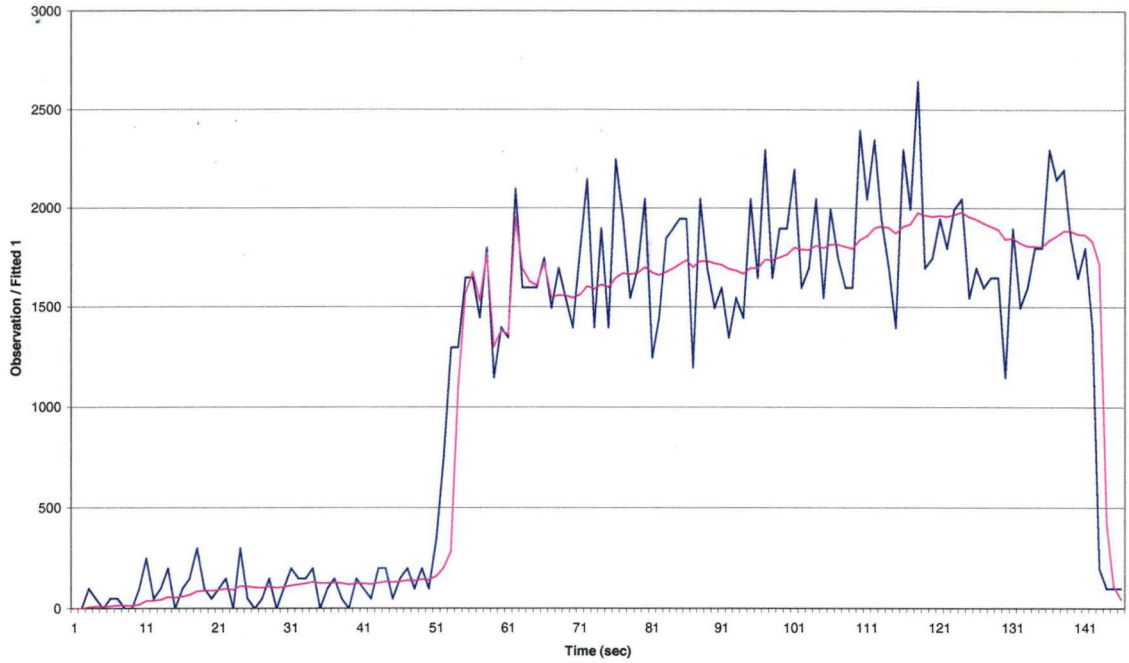


Fig 5.4 (3) : *HOLT Model Of Cu65, Sample 5. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 5 in element Cu65, whereas the smooth curve is the HOLT model with T = 9.0 (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

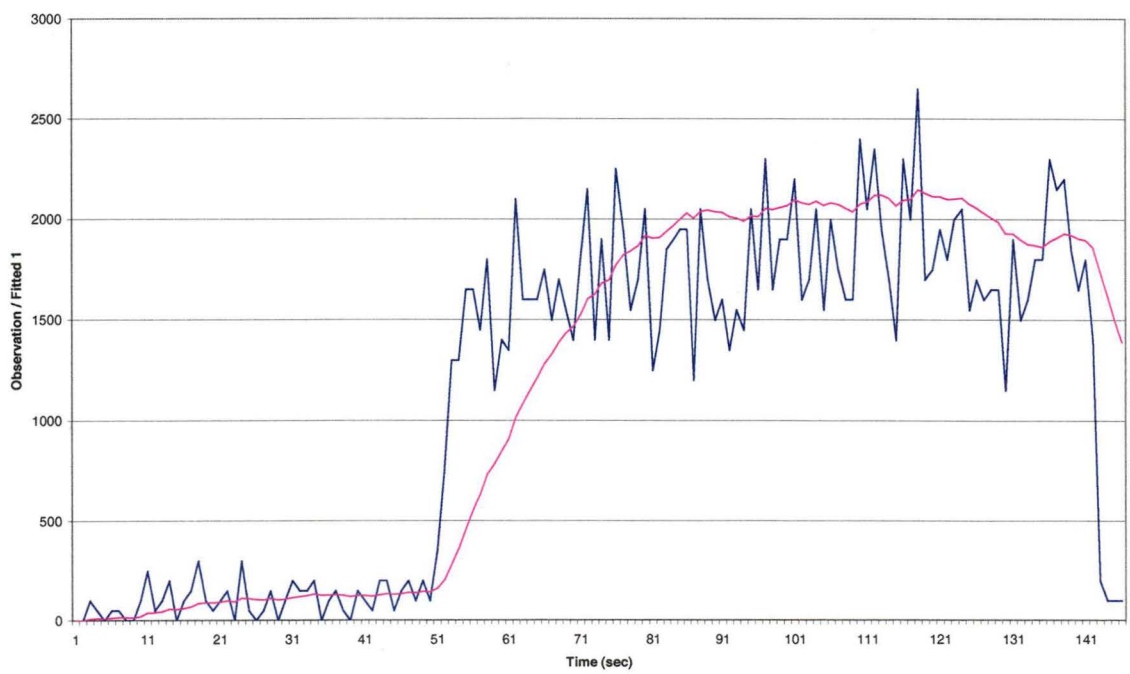


Fig 5.4 (4) : *HOLT Model Of Cu65, Sample 5. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 5 in element Cu65, whereas the smooth curve is the HOLT model with T = 50.0 (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

(5.5) Chapter Conclusions

In this chapter, the linear Holt exponential smoothing model is chosen to fit to the tab-chart. However, this model cannot be applied directly for tab-chart and there is some modification. The new modified model is called HOLT model. This model has the two pairs of parameters (i.e. (α_1,β_1) and (α_2,β_2)), which is to adjust the fitting of the model to the tab-chart, and has another parameter (i.e. T), which is to control of choosing the parameters (α_1,β_1) or (α_2,β_2). By experiment experience, the parameter T is more influential and important than the parameters (α_1,β_1) and (α_2,β_2). When T value is small, the HOLT model is spiky. As the value T increases, the model will becomes more and more smooth. However, the large T value will cause the large deviation from the tab-chart.

Chapter 6 Classification Of The Tab-Chart Of Secondary Standard Sample

(6.1) The Work Of This Chapter

In Chapter 5, the HOLT model was introduced and fitted to the tab-charts by varying the parameters (i.e. α and β) and threshold value (i.e. T). In this chapter, there are two goals and themes. The first goal is to devise a classification method to help the HOLT model to divide the ablation signals of the tab-charts of primary and secondary samples into different portions or groups (i.e. background, jump, plateau and drop). Then graphs of trend estimate versus smooth estimate in HOLT model are plotted. These graphs show us that the different portions of the ablation signals form different clusters. The second goal is to determine the classification rules that best classify the tab-chart. Since the signal clusters can be demarked by polygons, the classification rules are derived from the equations for the polygons. The effectiveness of these rules is evaluated by applying them to the primary and secondary samples.

(6.2) Classification Of Tab-Chart

(6.2.1) Description Of The Classification Method

To begin, the data value (i.e. the signal intensity) in the tab-chart is labelled one (the label given to the background phase). If the HOLT model changes the parameter sets, the label value of the next data will increase by one. Otherwise it retains the old label value. The data having value two are classified as "jump". Label value three is interpreted as a plateau. The data, having the maximum value and are not labelled three, and classified as "drop", whereas the intermediate value between two and maximum value are classified as "plateau". This algorithm is set out in a flowchart in Fig 6.2 (1).

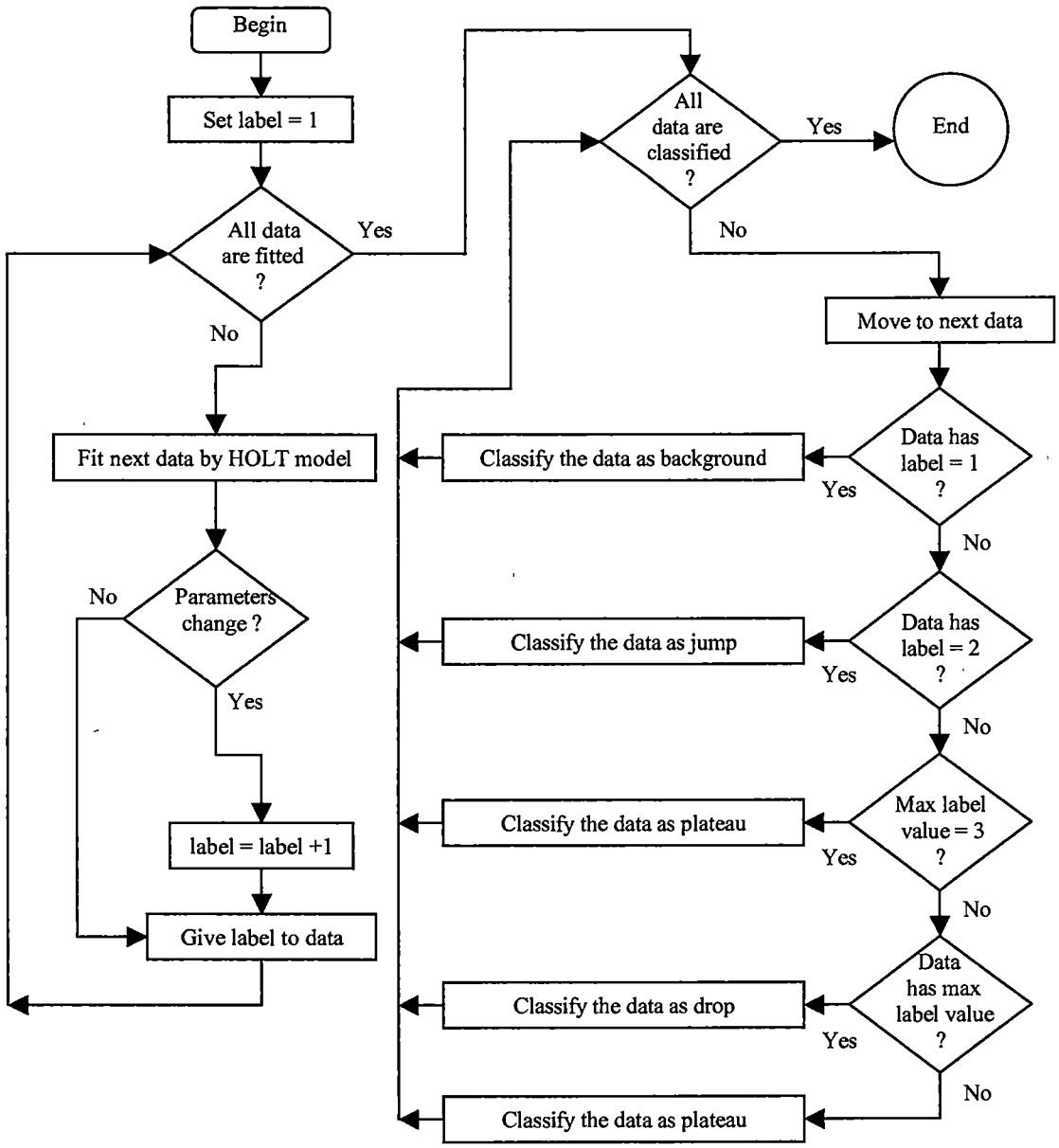


Fig 6.2 (1) : Classification Of The Tab-Chart

(6.2.2) Statistical Summaries After Applying The Classification Method To Tab-Chart

After the positions (i.e. starting points and ending points) of the background and plateau in a tab-chart are determined by the classification method, the statistical summaries of the background and plateau (i.e. the average level and standard deviation of the count per second over background and plateau) can be obtained immediately. The formulas or equations of calculating the statistical summaries in this method are the same as that in Intuitive method, but with different definitions in the notations (see section 4.3.5). The new definitions are listed in the following.

(a) x_k = the fitted values (i.e. $\hat{Y}_t(0)$ or $\hat{Y}_{t-1}(1)$) calculated by the HOLT model

(b) Φ = observations in the tab-chart are classified as background by *classification method*

(c) Ω = observations in the tab-chart are classified as plateau by *classification method*

(d) n_{BG} = total number of observations is classified as background by *classification method*

(e) n_{SG} = total number of observations is classified as plateau by *classification method*

The mean (average level) of plateau of samples 2 and 44 are provided in the following graphs (see Fig 6.2 (2) and Fig 6.2 (3)). An example of calculating the statistical summaries is shown in the Appendix (see Appendix F). The statistical summaries of elements in all samples are also provided in the Appendix (see Appendix H and I).

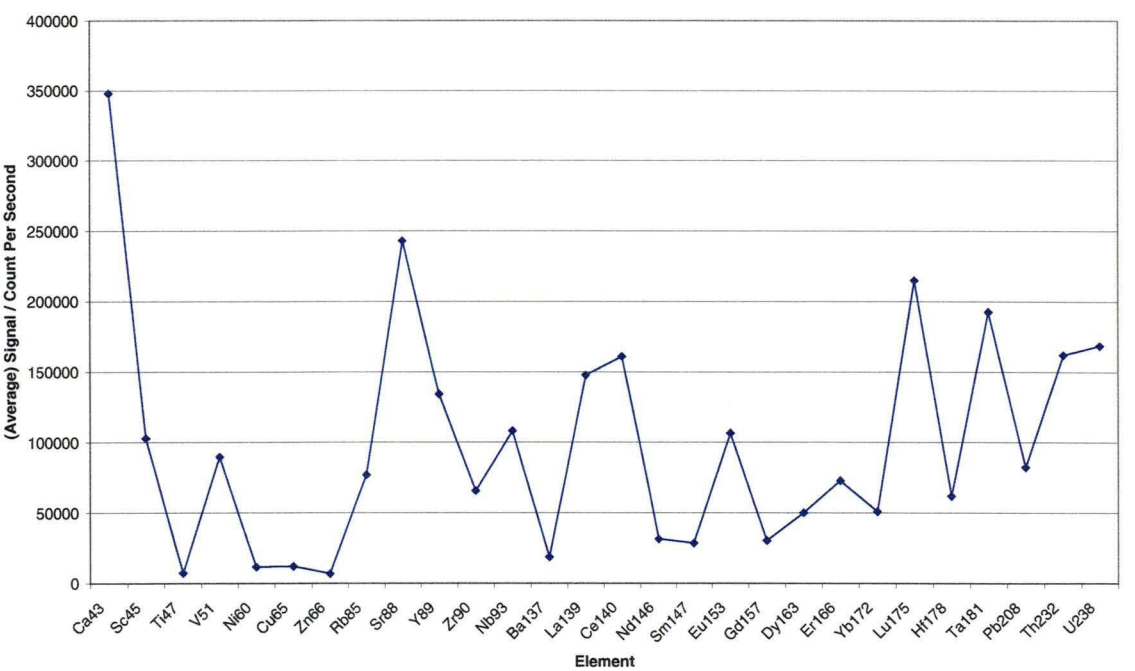


Fig 6.2 (2) : *Average Level Of Plateau (Sample 2). The average signal is calculated by the fitted value* $\hat{Y}_t(0) = L_t$

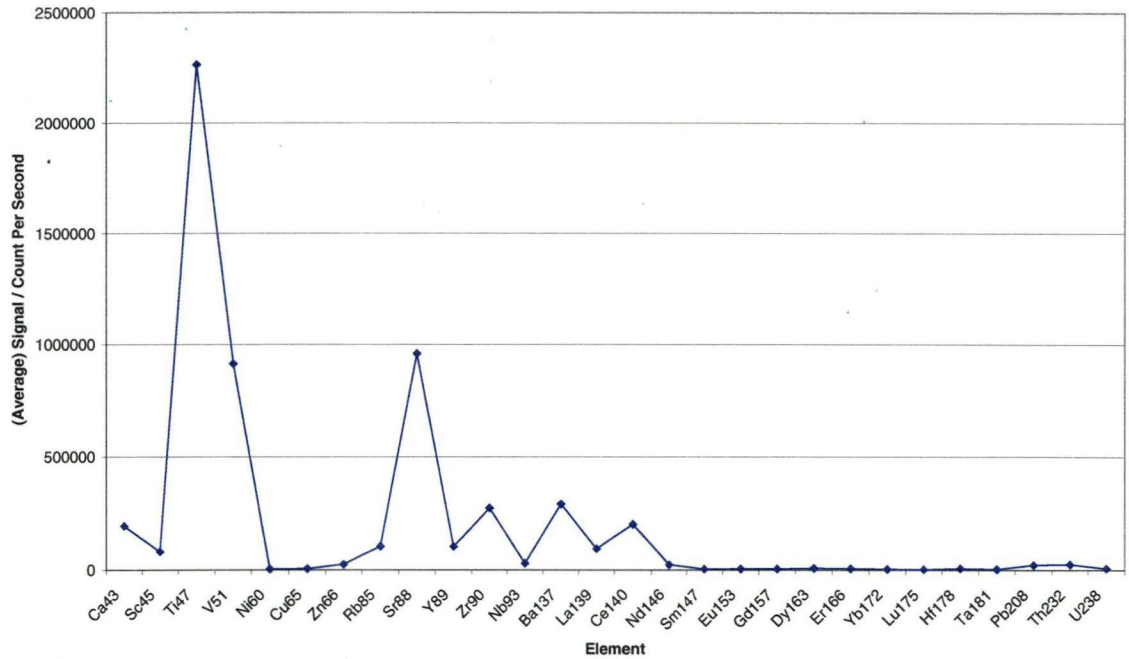


Fig 6.2 (3) : *Average Level Of Plateau (Sample 44). The average signal is calculated by the fitted value*

$$\hat{Y}_t(0) = L_t$$

Actually, the samples 1, 2, 33, 34, 65 and 66 are primary standards, which are used to correct the machine drift. The other samples are secondary standards (see Chapter 3). From the above two graphs, the average signals of elements in the primary samples are evenly distributed except Ca43. The secondary samples contain more light elements than heavy elements; the average signal of element decreases as the weight of the element increases.

When studying the table of average level (mean) of background in the appendix (see Appendix H), it is surprising that a lot of mean values in the appendix are much higher than the actual mean of background of the tab-chart. The cause is possibly the lag-behind effect of the classification method. The early stage of jump or plateau in some tab-charts is classified as background. In the sense of data analysis, these observations are the outliers and should be deleted before applying the equations of statistical summaries (i.e. pre-process the data). This problem is not hard to be solved because there are methods that can be found in the data analysis textbooks.

(6.2.3) Statistical Summaries Of Concentration After Applying The Classification Method

As for the statistical summaries of the background and plateau after applying the classification method (see section 6.2.2), the formulas or equations of calculating the mean and standard deviation of the element concentration after applying the classification method is the same as that of Intuitive method (see section 4.3.5), but use the notations which are defined in the section 6.2.2.

(6.2.4) ***Analysis And Result Of The Classification Method***

After classifying all samples by the above method, graphs of "trend" against "smooth" in the HOLT model are plotted. By looking into these graphs, the following observations are made. The data (i.e. signal intensity) which are classified as background, have small smooth and trend estimate values. The data, which are classified as jump, have large values of trend and smooth estimates. For the data of plateau, smooth is much larger than trend. For the data of drop, trends have large negative values.

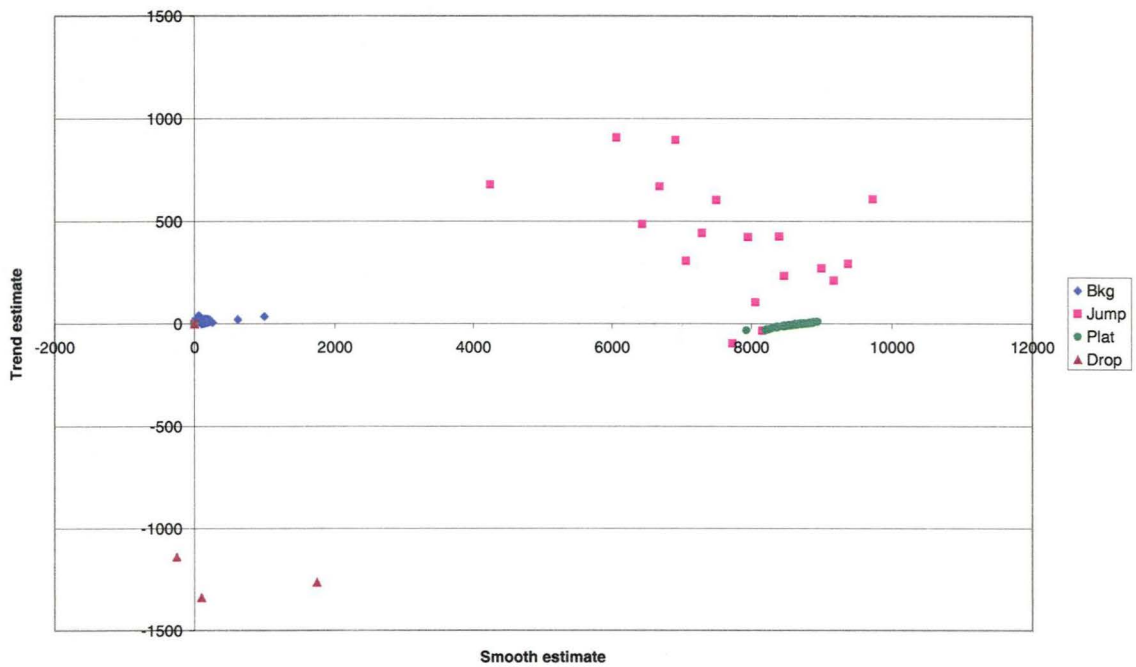


Fig 6.2 (4) : *Plot Of Smooth Versus Trend (Nb93, Sample 3). By classification method, blue diamond represents background, rectangle represents jump, green circle represents plateau and triangle represents drop*

By comparing the graph of smooth versus trend to the tab-chart in each sample, a phenomenon is observed and drawn. When the tab-chart is in the phase of background, the HOLT model will produce small smooth and trend and close to the origin; the points (smooth, trend) concentrate at the origin. In some samples, the trend is high relative to smooth at the initial because the HOLT model needs time to track the background data (i.e. transient state). When the tab-chart enters the phase of jump, the points start to move with positive slope; both smooth and trend increase significantly but the magnitude depends on the level of plateau in the tab-chart. Then the points fluctuate because the tab-chart enters from the phase of jump to the phase of plateau. Eventually the points move close to the horizontal axis because the HOLT model has detected the phase of plateau. Finally, when the tab-chart reaches the phase of drop, the points will move back to origin with clockwise movement.

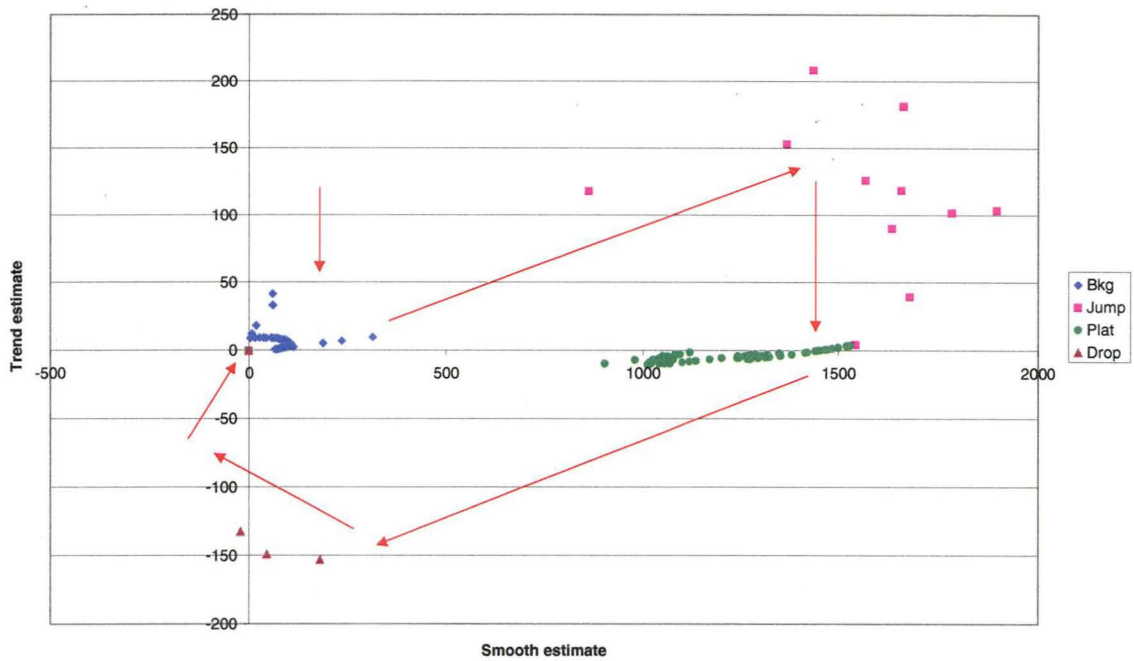


Fig 6.2 (5) : *Movement Of Point (Smooth, Trend) In the Span Of Tab-Chart (Zn66, Sample 23). By classification method, blue diamond represents background, rectangle represents jump, green circle represents plateau and triangle represents drop; the red line shows the change of parameters along the tab-chart*

(6.2.5) Defect Of The Classification Method

The classification method has defect such that it will produce inadequate results in some situations; it may misclassify most parts of tab-chart in some samples. The first defective example (i.e. Fig 6.2 (6) and Fig 6.2 (7)) is when the level of background and plateau are too close, whereas the second defective example (i.e. Fig 6.2 (8)) reveals the case where the HOLT model does not clearly distinguish between the plateau phase and drop phase in a tab-chart.

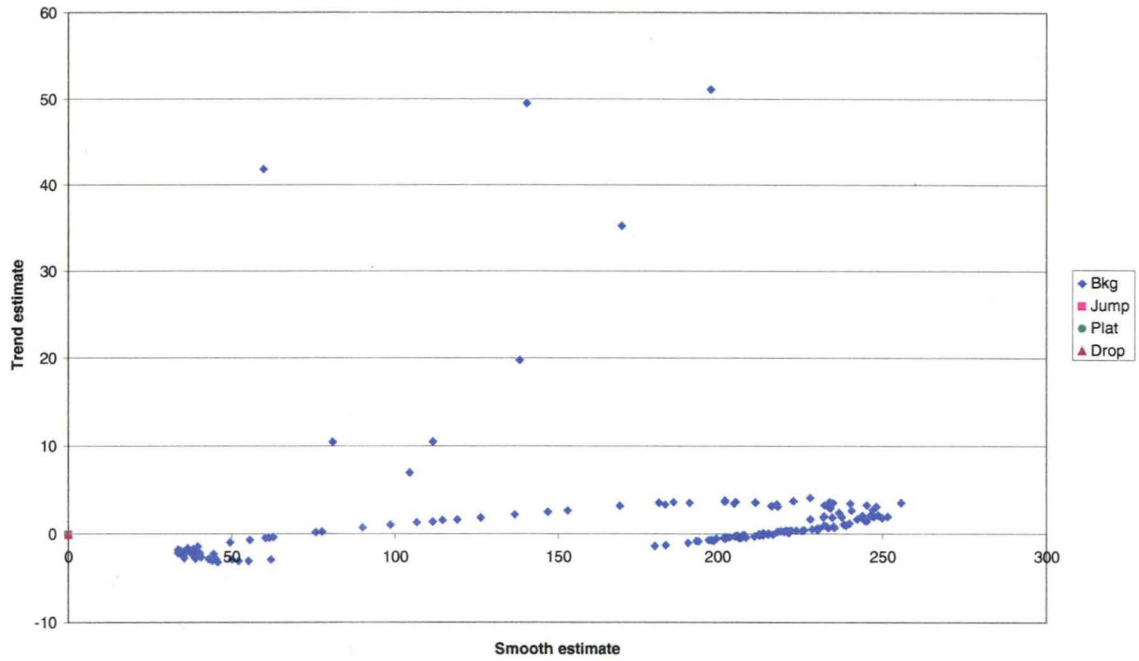


Fig 6.2 (6) : *Defective Example 1 (Cu65, Sample 16). By the classification method, the blue diamond represents that all data of tab-chart are classified as background*

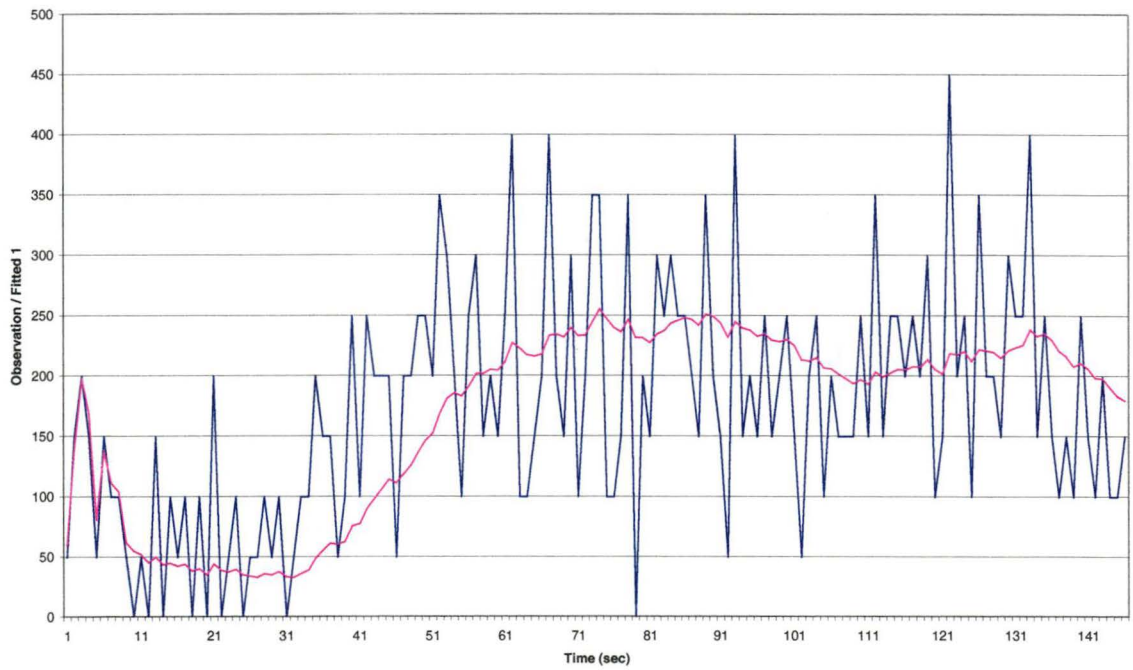


Fig 6.2 (7) : *Defective Example 1 (Cu65, Sample 16). The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 16 in element Cu65, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

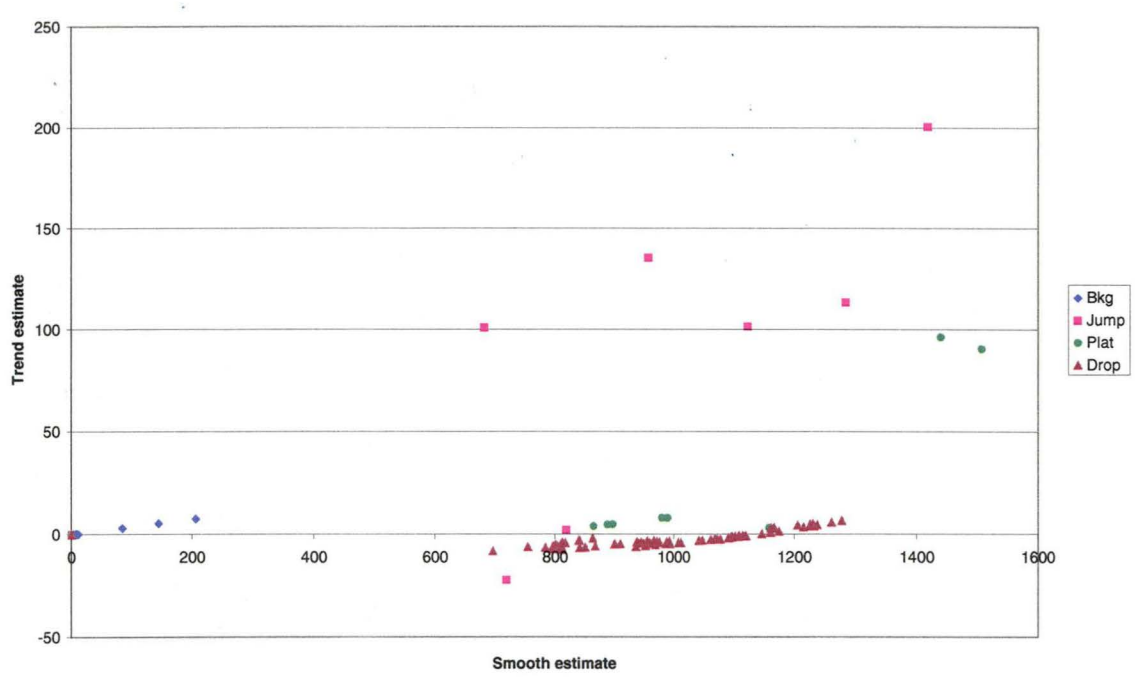


Fig 6.2 (8) : *<u>Defective Example 2 (Eu153, Sample 13)</u>. By classification method, blue diamond represents background, rectangle represents jump, green circle represents plateau and triangle represents drop. The classification method is slow to respond to the phase change from plateau to drop*

However, the number of misclassifications (i.e. tab-charts are defective) is not large after applying the classification method to all the tab-charts. Thus these misclassifications can be ignored in the further analysis. The proposed classification method is a way of helping us to find out the general rules to classify tab-charts automatically.

(6.3) Automatic Rules Of Classifying Tab-Chart

(6.3.1) Finding The General Rules For Classification

After putting all trend-smooth plots of all samples in each element together (i.e. all tab-charts of each element), it is found that the background can be approximately confined by a rectangle. The size of the rectangle is defined such that the rectangle confines the cluster of points near the origin. The jump, plateau and drop can be approximately confined by different triangles, with one of the vertices at the origin. Moreover, the triangle of plateau lies inside the triangle of jump.

The trend-smooth plots of all samples in Pb208 are listed in the following (Fig 6.3 (1) to Fig 6.3 (5)). In the first diagrams (i.e. the trend-smooth plots over background phase), some points are outside the rectangle because they should be classified as jump. The reason is that their smooth estimate and trend estimate obviously start to increase significantly outside the rectangle, so the fitted values $\hat{Y}_t(0)$ and $\hat{Y}_{t-1}(1)$ start to increase significantly (see sections 5.3 and 5.4.2). In the fourth diagrams (i.e. the trend-smooth over drop phase), the points close to horizontal axis should be excluded from the triangle because they should be classified plateau. The reason is that their smooth estimates are large and their trend estimates are close to zero (see section 5.3).

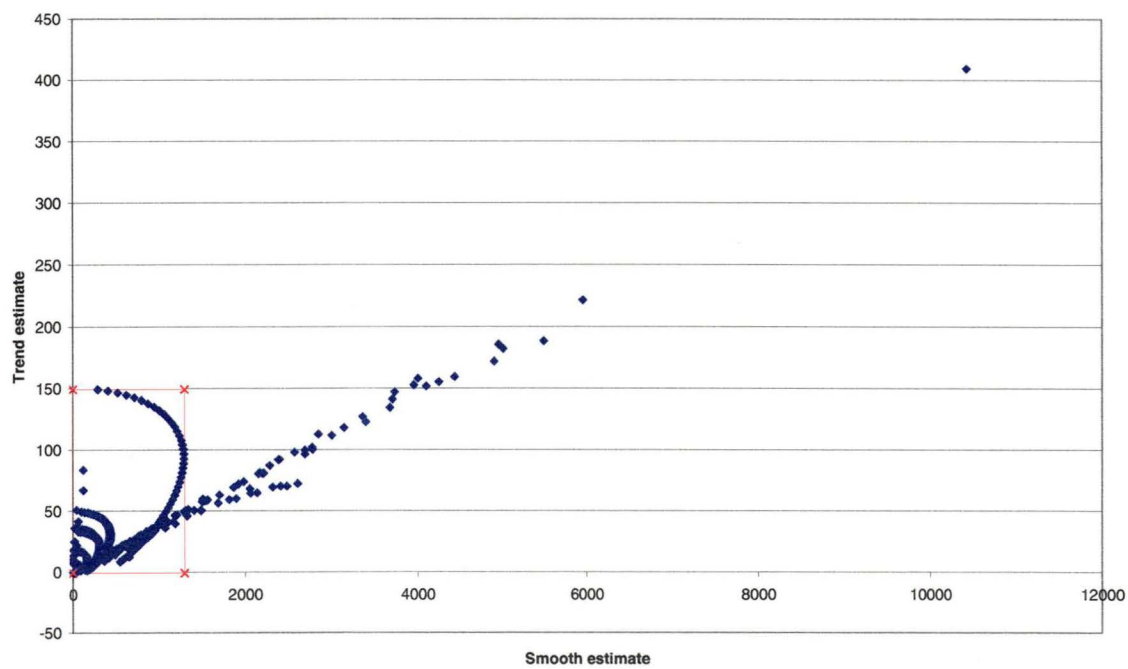


Fig 6.3 (1) : *Plot Of Smooth Versus Trend For Background (Pb208, All Samples). Blue diamond plot symbol represents background. The red rectangle is defined to classify whether the data is background or not*

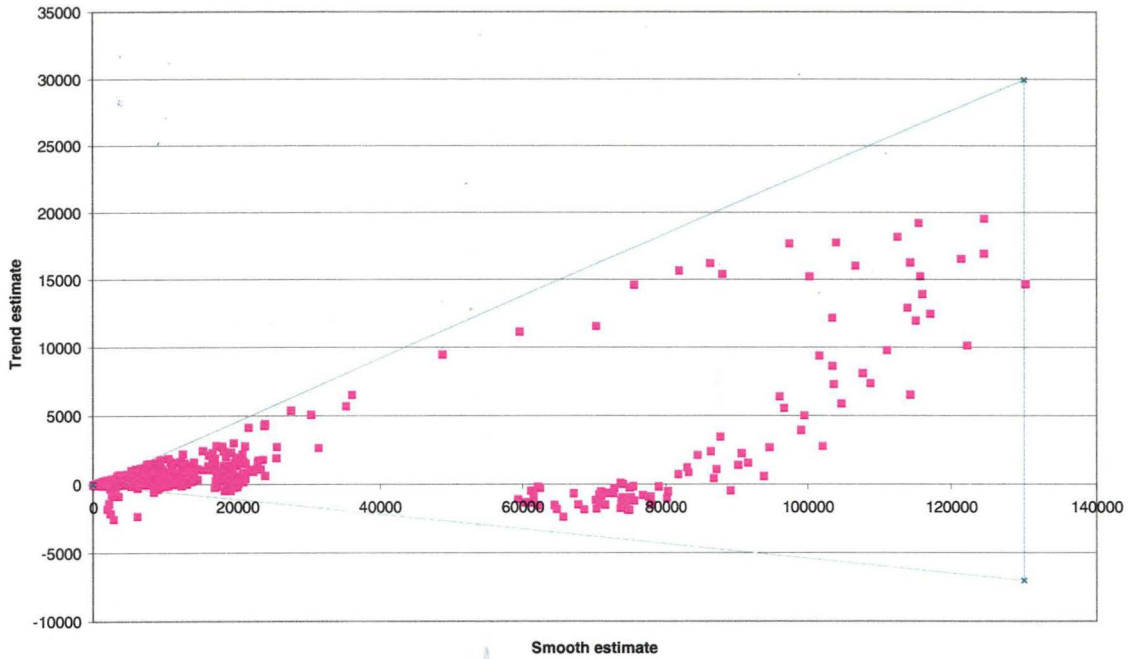


Fig 6.3 (2) : *Plot Of Smooth Versus Trend For Jump (Pb208, All Samples). Rectangle plot symbol represents jump. The green triangle is defined to classify whether the data is jump or not*

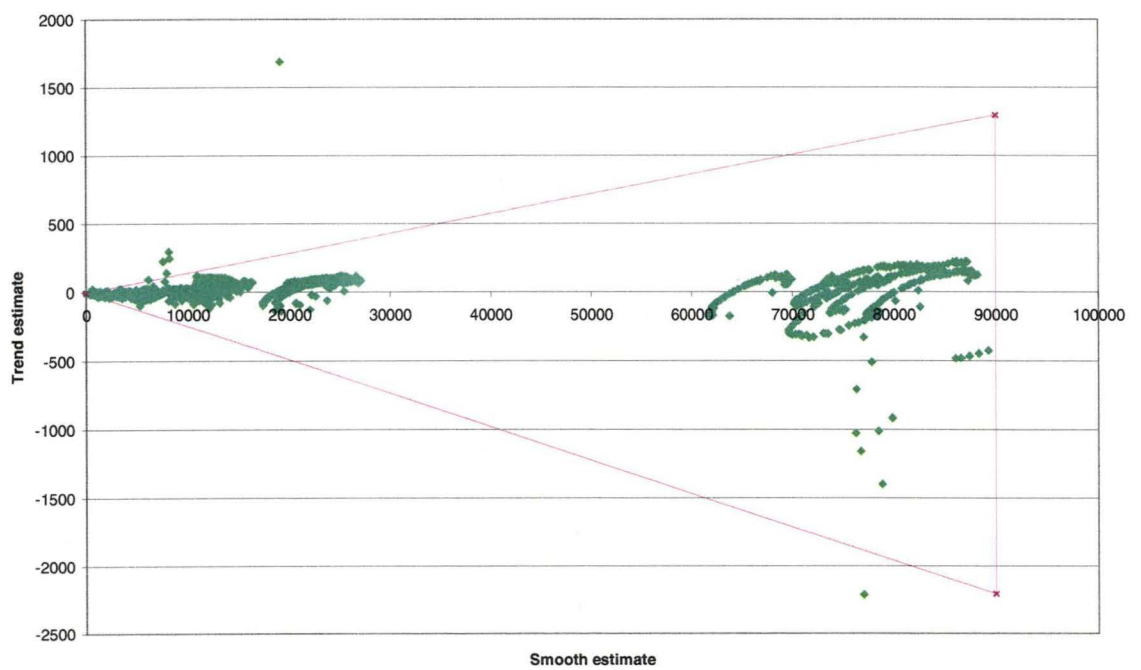


Fig 6.3 (3) : *Plot Of Smooth Versus Trend For Plateau (Pb208, All Samples). Green circle plot symbol represents plateau. The purple triangle is defined to classify whether the data is plateau or not*

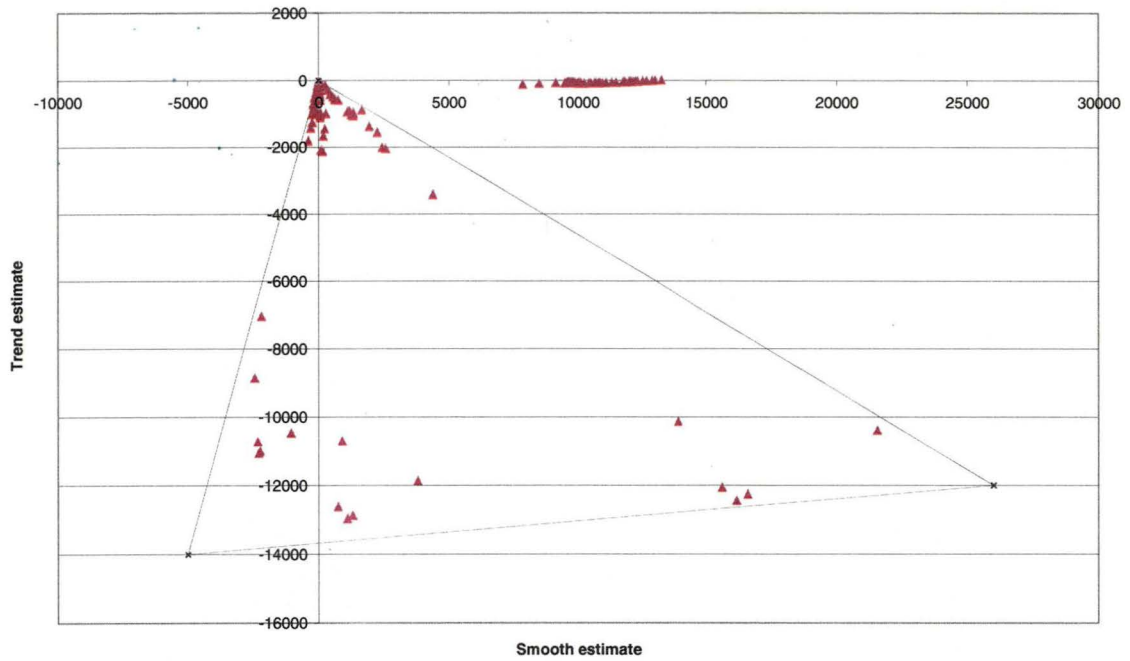


Fig 6.3 (4) : *Plot Of Smooth Versus Trend For Drop (Pb208, All Samples). Triangle plot symbol represents drop. The black triangle is defined to classify whether the data is drop or not*

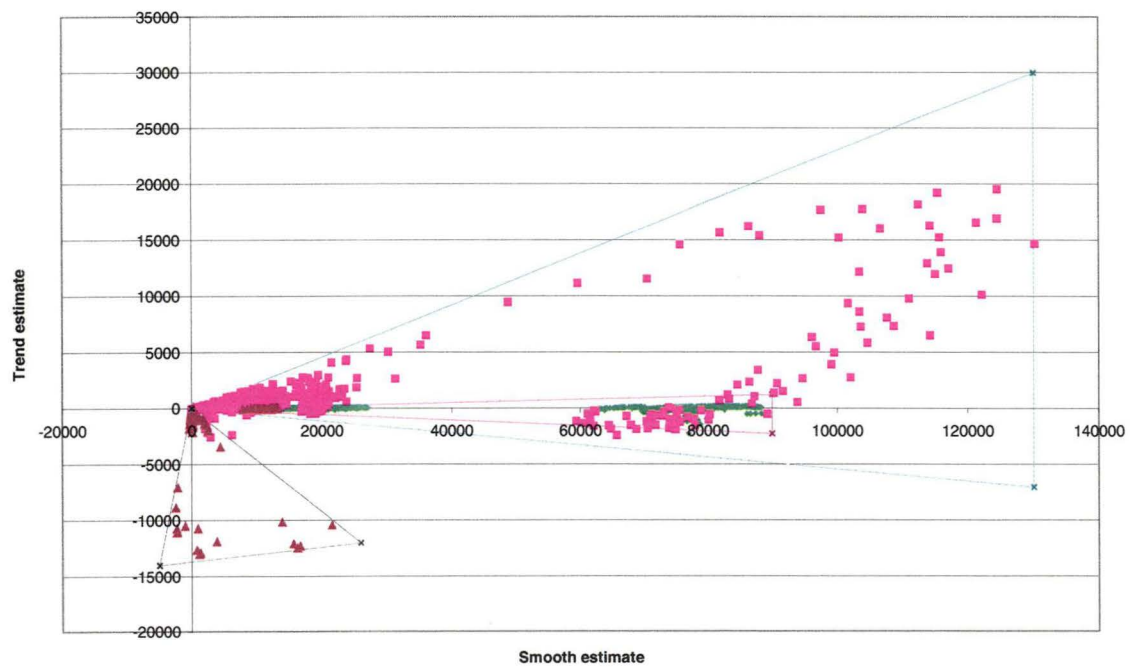


Fig 6.3 (5) : *Plot Of Smooth Versus Trend (Pb208, All Samples). Plot Symbol : Blue diamond represents background, rectangle represents jump, green circle represents plateau and triangle represents drop. Lines : The triangles and rectangle are defined in order to confine different phase under the classification method*

After studying the equations of the confined regions in all elements, the approximated classification rules are drawn and listed as the following flowchart (Fig 6.3 (6)).

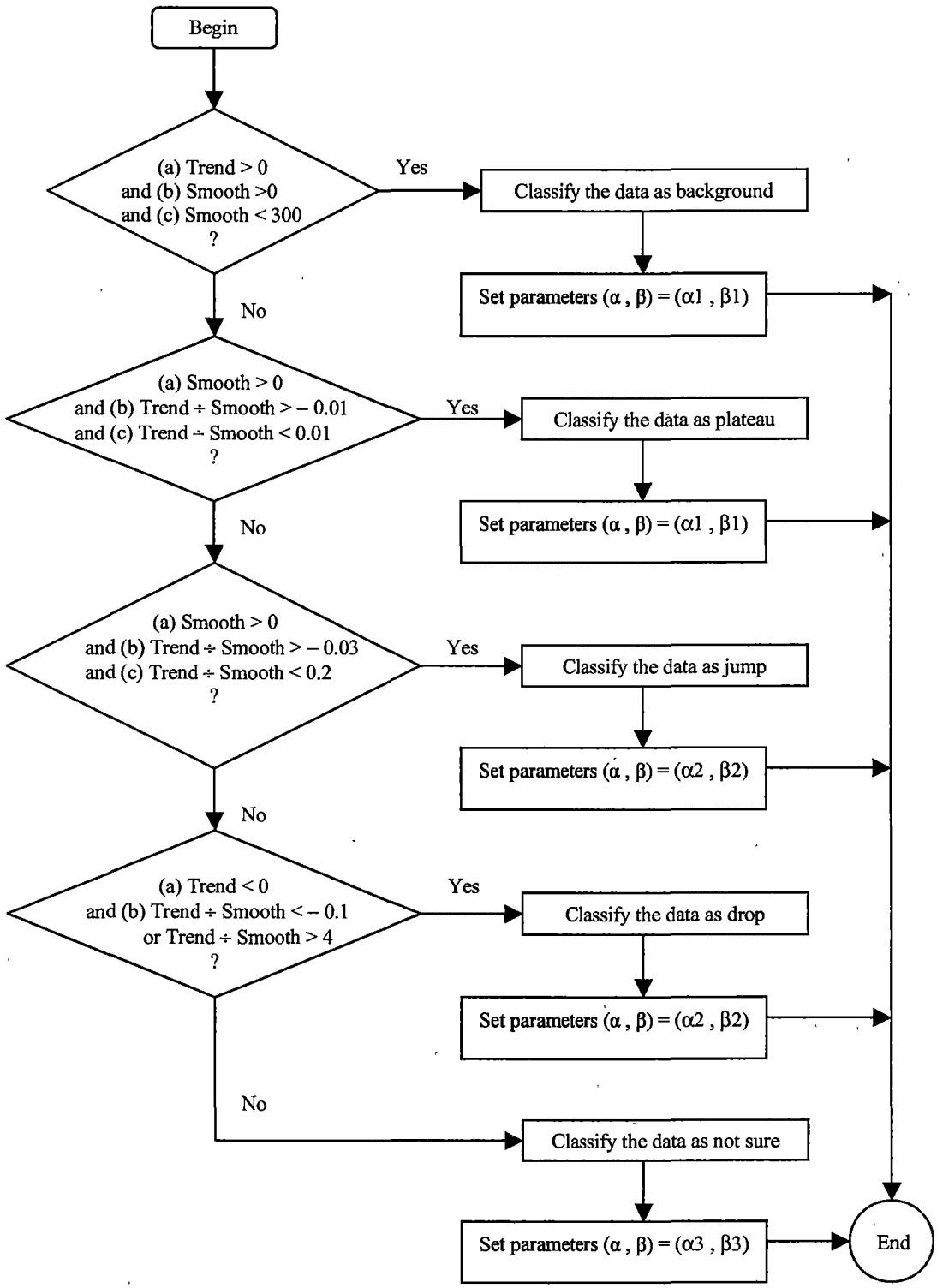


Fig 6.3 (6) : *Flowchart Of The Rule Of Classifying Tab-Chart And Parameters Setting*

(6.3.2) ***Statistical Summaries After Applying The Classification Rules To Tab-Chart***

As for the classification method, the statistical summaries of the background and plateau can be obtained at once by the formulas after applying the classification rules to the tab-chart. The formulas or equations of calculating the statistical summaries in classification rules are the same as that in Intuitive method, but with different definitions in the notations (see section 4.3.5). The new definitions are listed in the following. Besides, an example of calculating the statistical summaries after applying classification rules is shown in the Appendix (see Appendix G).

(a) x_k = the fitted values (i.e. $\hat{Y}_t(0)$ or $\hat{Y}_{t-1}(1)$) calculated by the HOLT model

(b) Φ = observations in the tab-chart are classified as background by *classification rule*

(c) Ω = observations in the tab-chart are classified as plateau by *classification rule*

(d) n_{BG} = total number of observations is classified as background by *classification rule*

(e) n_{SG} = total number of observations is classified as plateau by *classification rule*

(6.3.3) ***Statistical Summaries Of Concentration After Applying The Classification Rules***

As for the statistical summaries of the background and plateau after applying the classification rules (see section 6.3.2), the formulas or equations of calculating the mean and standard deviation of the element concentration after applying the classification method is the same as that of Intuitive method (see section 4.3.5), but use the notations which are defined in the section 6.3.2.

(6.3.4) ***The Choice Of Parameters In The Classification Rules***

Before discussing our choice of parameters, two extreme cases of parameters are investigated and studied. The first case is to set all parameters to zero (i.e. $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0$). The fitted value of the HOLT model becomes a straight line or a zero horizontal line in some cases. It depends on the initial value of the tab-chart (see section 5.3 and Fig 6.3 (7)). The second extreme case is to set all parameters to one (i.e. $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1$). If the fitted value L_t is used, the curve of fitted value coincide the tab-chart. If the fitted value $L_{t-1} + T_{t-1}$ is used, the curve of fitted value is much spiky than the tab-chart (see Fig 6.3 (8)). By ignoring the effect on the result of classification rules, the HOLT model loses the power of data reduction (i.e. remove the noise and variation from the tab-chart) in these scenarios. The researchers cannot gain any information of the background and plateau (eg. trend) via the model after the tab-chart is partitioned by classification rules. There is no difference between the current change point detection (or time series segmentation) techniques and HOLT model (see section 1.11). The quantitative analysis on the data reduction of the model will be discussed in the next chapter (see section 7.3.5).

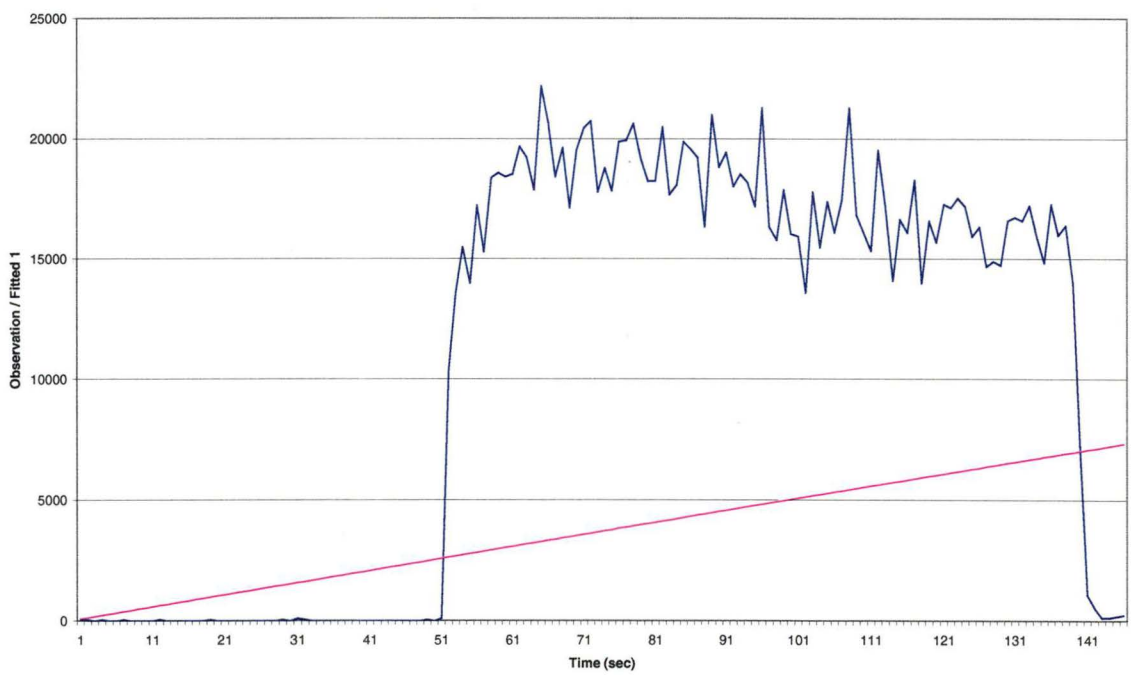


Fig 6.3 (7) : *All Parameters Are Set To Be Zero (Ce140, Sample 10). The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 10 in element Ce140, whereas the straight line is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

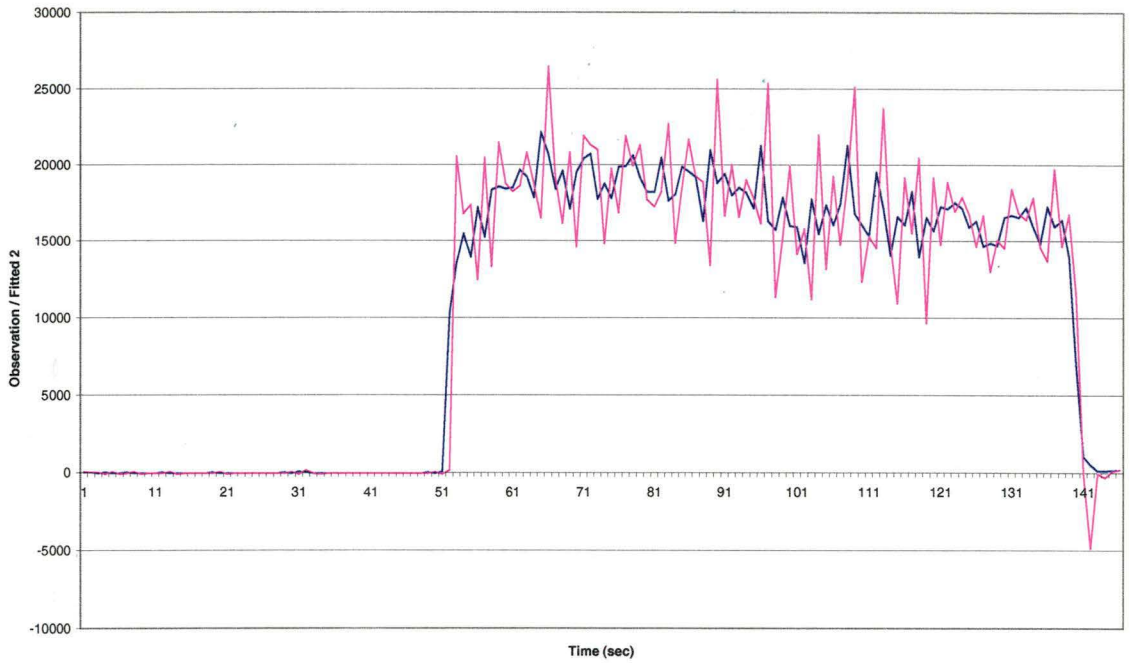


Fig 6.3 (8) : *<u>All Parameters Are Set To Be One (Ce140, Sample 10)</u>. The blue spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 10 in element Ce140, whereas the pink spiky curve is the HOLT model (i.e. the fitted value (b) of modified linear holt exponential smoothing)*

In the classification rules, the values of parameters are shown in the following table (see Fig 6.3 (9)). There may be a question mark here why these specific values are chosen. Actually, an extensive experiment should be done to find out the best parameters in various scenarios. In my opinion, I think that the classification rules have greater impact on the performance than that of parameters. This is an interesting problem for the further studies. Back to the table, the values of parameters in jump and drop are larger than that in background and plateau. When the jump or drop occurs in the tab-chart, it implies that there is a great value change over this period. In order to keep track of this change, more weight should put on the recent and current observations, Y_t (see section 5.3). On the other hand, smaller parameter values are used in the background and plateau. More weight is put on the history of the tab-chart in order to keep track or the trend of the background and plateau (see section 5.3).

Fig 6.3 (9) : *<u>The Values Of Parameters In The Classification Rules</u>*

Phase	*Parameter α*	*Parameter β*
Background	0.07	0.04
Jump	0.80	0.20
Plateau	0.07	0.04
Drop	0.80	0.20
Not Sure	0.50	0.50

(6.3.5) ***Guidelines To Evaluate The Classification Rules***

The classification rules will be applied back to the total 1848 tab-charts of primary and secondary standard samples. The performance of the classification rules is evaluated by studying each tab-chart. When evaluating the performance, the raw data and tab-chart of the sample and the results of the classification rules will be studied together. Although the evaluation process involves the personal judgement, some subjective guidelines are made to decide whether the rules are successful on the samples or not. In some borderline cases (i.e. samples), different persons possibly have different conclusions. The tab-chart of the standard sample has the following shape. The first stage is background, the second stage is jump, the third stage is plateau and the last stage is drop; this shape or pattern is used to test the performance of the classification rules. The test guidelines are outlined as the following paragraphs.

If the classification rules can give the result similar to above pattern, the rules are absolutely satisfactory or successful in this sample. For example, samples 1, 3 and 14 in Ta181 (see Appendix D). In the following graph, the classification rules can produce good performance on Ta181, sample 3 (see Fig 6.3 (10)).

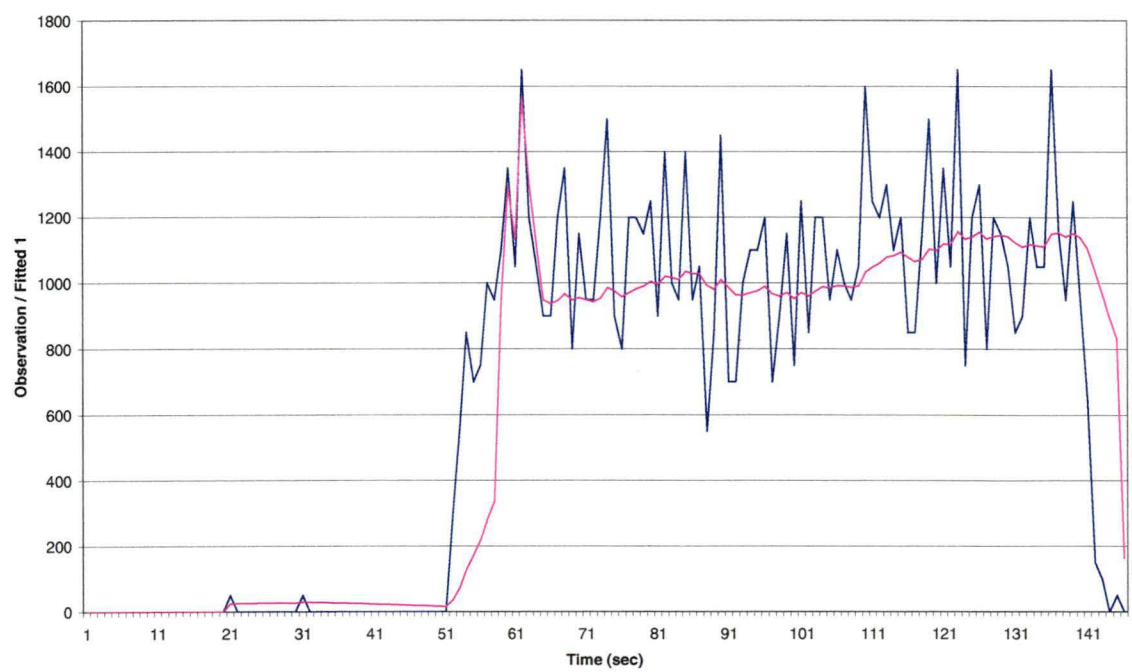


Fig 6.3 (10) : *The Classification Rules Can Work In This Sample (Ta181, Sample 3). The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 3 in element Ta181, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

In some samples, the performance of the rules is also fairly satisfactory although there are some unclear patterns in the result. It is because the rules still provide us a segment of background and then a segment of plateau. For example, samples 7 and 35 in Ta181 (see Appendix D). In the following graph, the classification rules can produce fair satisfactory performance on Ta181, sample 35 (see Fig 6.3 (11)).

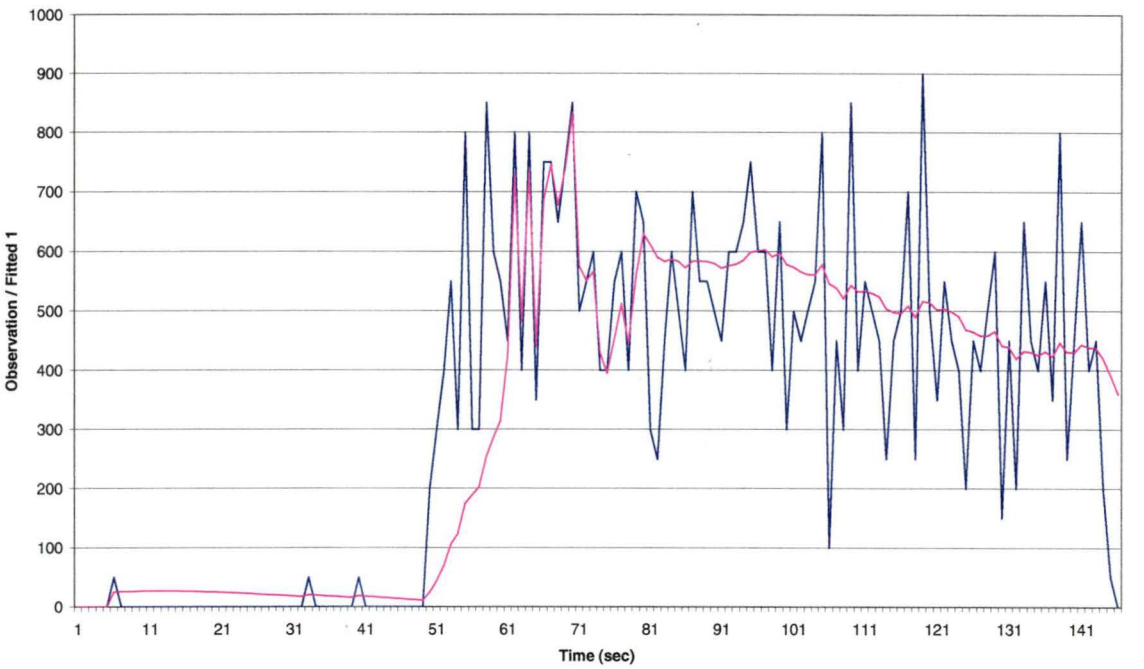


Fig 6.3 (11) : *The Classification Rules Can Work In This Sample (Ta181, Sample 35). The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 35 in element Ta181, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

If the classification rules cannot give the clear pattern in some part of the result, the rules are classified as unsatisfactory or unsuccessful in these samples. There are some "fluctuations" in the result to confuse the judgement. For example, samples 25 and 40 in Ta181 (see Appendix D). The classification rules obviously and absolutely fail in some samples. In other words, there are a lot of misclassifications in these samples. For example, samples 17, 22 and 52 in Ta181 (see Appendix D). In the following graphs, the classification rules fail and cannot work on Ta181, samples 17, 22 and 52 (see Fig 6.3 (12), (13) and (14)). The tab-charts of these three samples clearly show that there are background and plateau, however the rules are not capable of giving these results. There is an interesting point that the background and plateau of these tab-charts are very close and the jump is not obvious (see section 6.3.7).

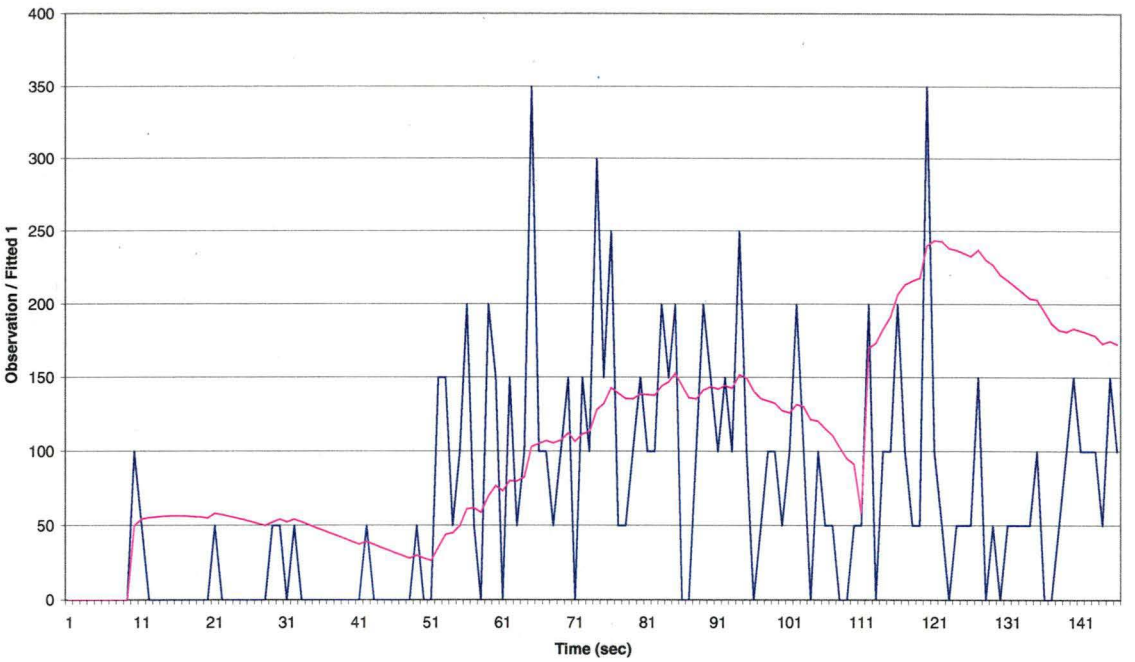


Fig 6.3 (12) : *<u>The Classification Rules Cannot Work In This Sample (Ta181, Sample 17)</u>. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 17 in element Ta181, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

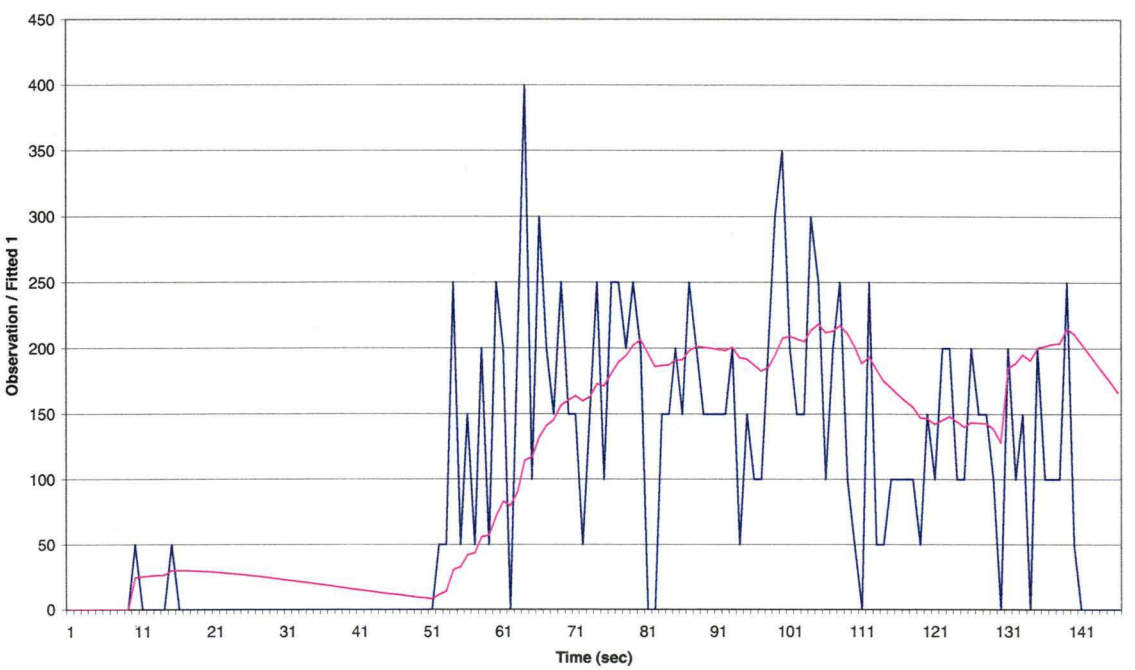


Fig 6.3 (13) : *<u>The Classification Rules Cannot Work In This Sample (Ta181, Sample 22)</u>. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 22 in element Ta181, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

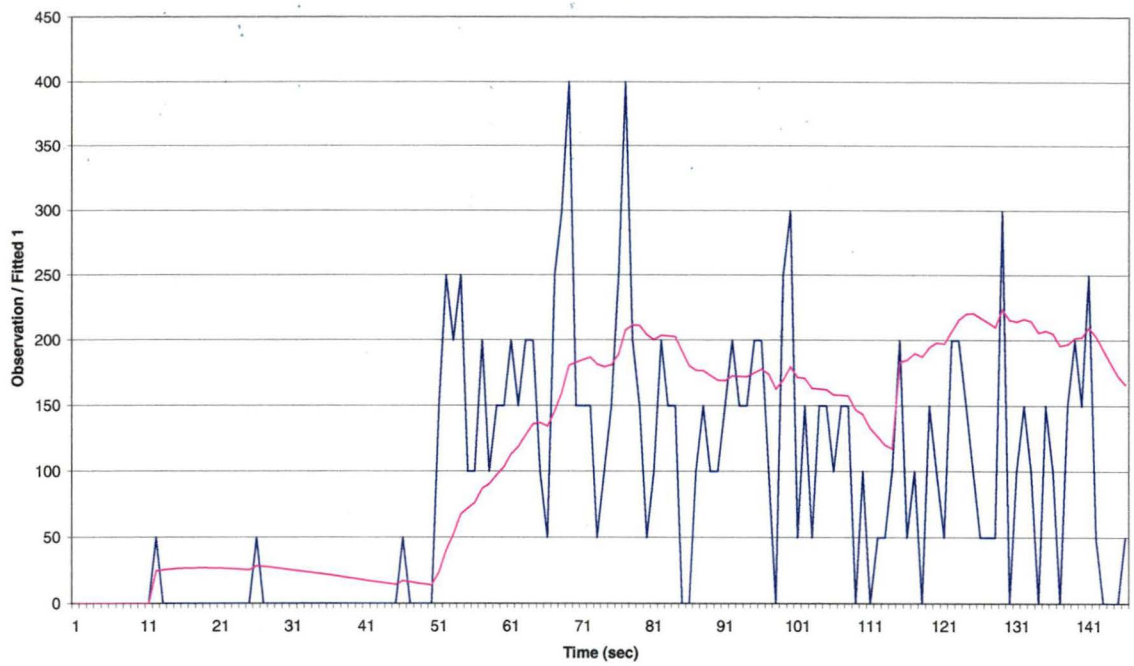


Fig 6.3 (14) *: The Classification Rules Cannot Work In This Sample (Ta181, Sample 52). The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 52 in element Ta181, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

I recognize there is some bias on the figures in the evaluation tables (see Appendix E). Although the guidelines are established in the previous section (see section 6.3.5), it is not easy to make the judgement in some borderline cases (i.e. samples). However, the figures are helpful and good references such that the weakness of the rules can be identified and discovered. The purpose of this thesis is to explore new model and method to apply to the tab-chart. There is a lot of room to refine and modify the classification rules in the future. Further work might include a structure and comparison with classifications by one or more human experts.

Before closing this section, there are several remarks for discussion. Firstly, since the HOLT model and rules need time to respond and detect the stage change of the tab-chart, it is impossible to pinpoint the start and end time of each stage. In other words, the change of results will lag behind the stage change of the tab-chart. Secondly, the period of drop stage in the tab-chart is very short. The performance evaluation will focus on the background, jump and plateau. Thirdly, element Ca43 is special and different from other elements. By comparing their raw data, Ca43 has two plateaus, whereas the other elements have one background and plateau.

(6.3.6) Performance Of The Classification Rules

The classification rules are applied back to the tab-charts of the standard samples. The classification rules have the lowest successful rate (i.e. performance) of about 45% in elements Cu65 and Dy163. On the contrary, the classification rules have the highest successful rate of above 80% in elements La139, Nb93, Nd146, Y89 and Zr90. For other elements, the successful rate ranges from 50% to 75% (see Appendix E). After excluding the samples of having close background and plateau, the successful rate of the classification rates will increase to 65% or above (see Appendix E). The guidelines to identify these specific samples (or tab-charts) will be explained at the next section (see section 6.3.8).

To sum up, the classification rules have good performance in tackling the samples, which have significant signal intensity and the background and plateau are not too close. On the contrary, the rules become much worse as the signal intensity (i.e. counts per second) get smaller and the background and plateau get closer. There are some suggestions that can make the rules more perfect. Firstly, the rules are relatively simple since it is in the exploration stage to use time series to segment the tab-chart (i.e. detection of change point). More complex rules should be devised to improve the performance. Secondly, the samples can be grouped into different classes and study each class separately. Then different classification rules are devised for different classes. Thirdly, other mathematical method can be incorporated into our current method.

(6.3.7) Discussions On The Tab-Chart Having Close Background And Plateau

Both classification method and rules have the worse performance on the tab-charts of having very close background and plateau. By studying these tab-charts (see Fig 6.2 (7), Fig 6.3 (12), Fig 6.3 (13) and Fig 6.3 (14)), the amount of the element is very small and some values in the plateau are zero. Besides, there is no obvious or too gradual jump in these tab-charts. In my opinion, there is a close relationship between "background and plateau is too close" and "the jump is too gradual".

(6.3.8) Guidelines To Identify The Tab-Chart Having Close Background And Plateau

The guidelines to identify these tab-chart are based on the following three criteria : (a) the signal intensity (i.e. counts per second) of the tab-chart is below a thousand; (b) there is no obvious jump existing between the background and plateau and (c) some parts of the plateau touch the x-axis (i.e. the value is or very close to zero) (see Fig 6.3 (12), (13), (14) and Appendix C).

(6.4) Chapter Conclusions

In this chapter, the classification method is devised after the HOLT model is well fitted to the tab-chart in previous chapter (see Chapter 5). The classification method has two functions in this thesis. Firstly, the classification rules are developed and devised from this method (see section 6.3). Secondly, this method will be used to analysis the data reduction and fitting error of the HOLT model in the next chapter (see Chapter 7).

The classification rules are used to partition the tab-chart into different portions in automotive way (see Fig 6.3 (6)). These rules work satisfactorily in more than 65% of the cases after excluding the tab-charts having close background and plateau (i.e. misclassification). After applying the rules to the tab-chart, this chapter also give the equations of calculating the statistical summaries of background and plateau. The example of these statistical summaries is given in the Appendix (see Appendix G). Besides, the equations of calculating the statistical summaries of the element concentration are also provided in this chapter.

The misclassification is a problem to both classification method and rules. This problem is that the classification and rules have the terrible performance in some tab-charts. In other words, there are a lot of observations in these tab-charts being wrongly classified. Although the misclassification affects the performance of classification method and rules, this problem give us the direction of improving the method and rules in the positive side.

There are several suggestions in the further studies. Firstly, the classification rules should be refined and modified to perform better performance because the rules are simple and rough. Secondly, an experiment should be done to investigate the influence of parameters on the classification's performance. Thirdly, statistical test or method should be done to remove the outliers before calculating the statistical summaries of background or plateau. Lastly, the classification rules are approximate and applied to cases of single jump and drop only. In the real cases, there are multiple jump and drops and plateau. The rules can be extended to the real cases (sample). The suggested algorithm is described as the following steps. The first data of the time series is classified as background and the classification rules are applied. When the jump and drop are detected by the classification rules, the new plateau caused by jump or drop will be the new background. The classification rules are applied to the time series until the last data (see Fig 6.4).

There is a remark before going to the next chapter. Although this thesis does not consider the real cases, the appropriate model and method are already explored and developed in this and previous chapters. The study of primary and secondary samples can build the good foundation and platform for studying real samples in the next stage. The knowledge of this thesis is still in exploration stage, the model and method are still in infant version. There are some inadequate places which more work is

needed on these places before moving to the real cases. For example, the problem of misclassification exists in the model and method and only one set of parameters is used in the analysis, etc. However, this thesis provides the worthwhile and promising direction for further studies. Besides, the knowledge of this thesis is not limited only on the composition investigation of rock sample. It can be generalized and applied to other areas (i.e. charts in other problem). The real cases can also be simulated in the computer for further studies and analysis even though the raw data is not available.

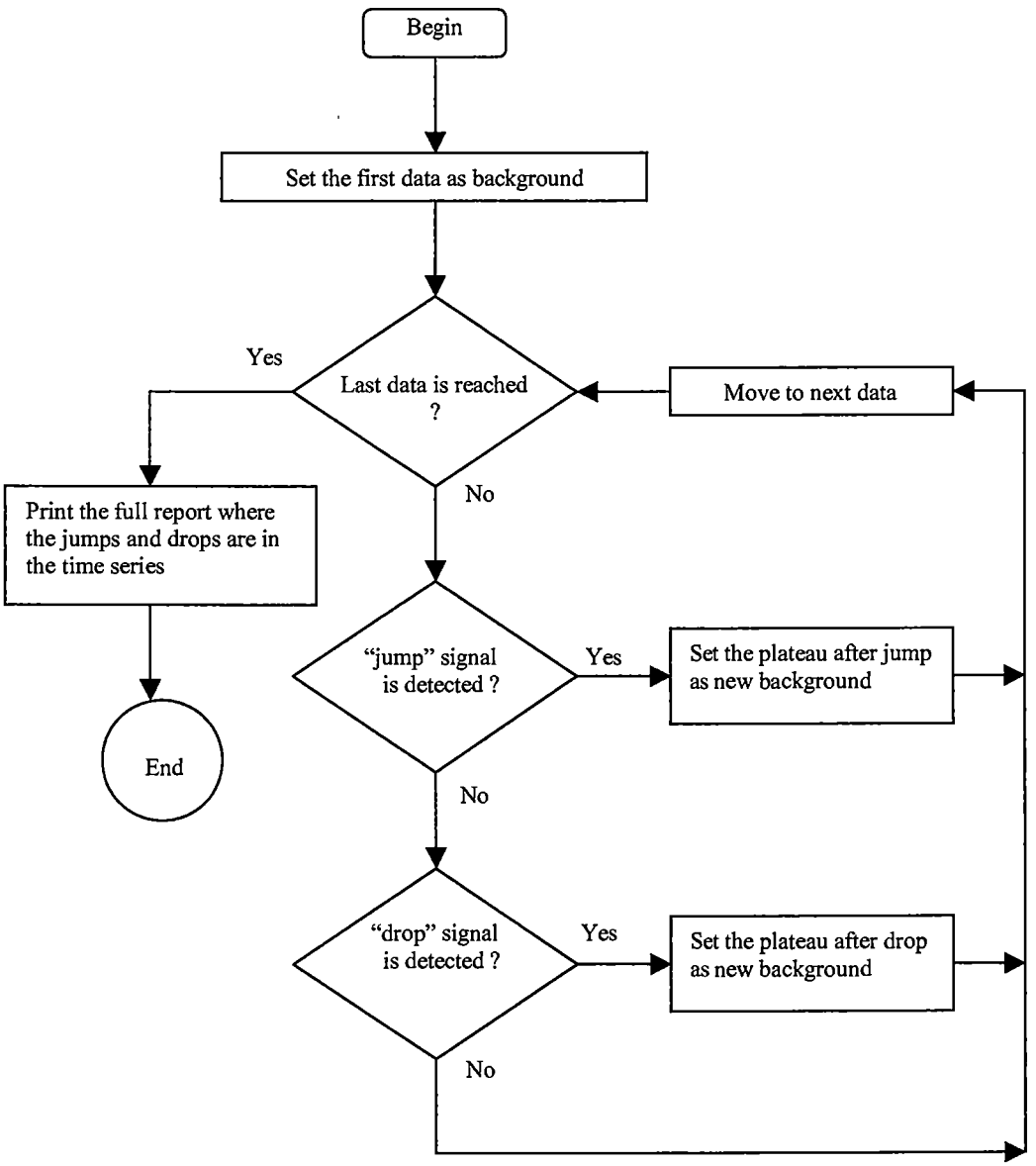


Fig 6.4 : *Rough Idea Of Applying The Classification Rules To Real Cases*

Chapter 7 Intuitive Method And HOLT Model

(7.1) The Work Of This Chapter

The "plateau" is the most important part of the ablation signal for obtaining concentrations, so in this chapter the aim is to conduct a statistical analysis of the error and noise estimate in the model, for observations in the plateau region. We also compare the standard error of the HOLT model and intuitive methods. This potentially depends on the method of classifying the observations in the plateau. This method is already discussed in the previous chapter (see Chapter 6). In the primary and secondary samples, it is easy to distinguish the plateau regions either by visual inspection or by a rule.

In the first part of this chapter, the standard error of the HOLT model is compared to that of the intuitive method in order to understand about the data reduction of HOLT model (see section 7.3). In the second part, the fitting performance of HOLT model to the tab-chart will be investigated. The relationship between the standard deviation of fitting error, average level of plateau and mass number are found and the equation will be provided (see section 7.4). There are several remarks to be discussed before starting the work of the statistical analysis on the Intuitive method and HOLT model.

(7.2) Review Of Previous Knowledge

(7.2.1) Review Of Intuitive Method

Before moving to next section, the intuitive method is reviewed and revisited. In this method, the researchers partition the tab-chart into different stages (portions) on their own by their experience, knowledge and visual judgement. The *integration interval for background* is the background portion that is partitioned by intuitive method; whereas *the integration interval for plateau* is the plateau portion that is partitioned by intuitive method. After the tab-chart is divided, the statistical summaries of these portions are calculated from the actual observation values (signal) of the tab-chart (see sections 4.3.1 and 4.3.5).

(7.2.2) Review And Comparison Of Classification Method And Classification Rules

First of all, the classification method and classification rules are reviewed and revisited. Both classification method and rules are to fit the modified linear Holt exponential smoothing model (i.e. HOLT) to the tab-chart. However, classification method is to divide the tab-chart into different portions by the human adjustment of the parameters and threshold value (i.e. α, β and T) (see sections 5.4.3 and 6.2.1), whereas the classification rules are divide the tab-chart into different portions automatically by some rules (see section 6.3.1).

The performance of classification method is better than that of classification rules by comparing their misclassifications, classification results and response time of stage change. For misclassification, both classification method and rules have this problem (see sections 6.2.4 and 6.3.7). For classification results, the classification rules will produce more unclear patterns and more fluctuations in the classification result (see Appendix D). For response time of stage change, the response time of stage change in classification rules is slower than that in classification method. The above comparison results are due to the different mechanism of classification method and classification rules. The classification method and classification rules use different approach to segment the tab-chart (see Fig 6.2 (1) and Fig 6.3 (6)). It is obvious that the human visual judgement (i.e. classification method) is often superior to a set of rules (i.e. classification rules). Therefore, the statistical and quantitative analysis of this chapter will use the segmentation result of classification method only because the result of classification method will make the analysis more easily and more clear conclusions will be drawn. I admit that a comprehensive analysis should also use the result of classification rules. However, the finding and conclusions from the classification method can be projected and can also be applied to the classification rules.

(7.2.3) Summary Of Previous Knowledge

The section is to provide the summary and clear picture of intuitive method, classification method, classification rules and HOLT model which are already discussed in the previous chapter (see chapters 4, 5 and 6). The first diagram in the following is to illustrate the application of intuitive method to the tab-chart, whereas the second diagram in the following is to illustrate the application of classification method, classification rules and HOLT model to the tab-chart.

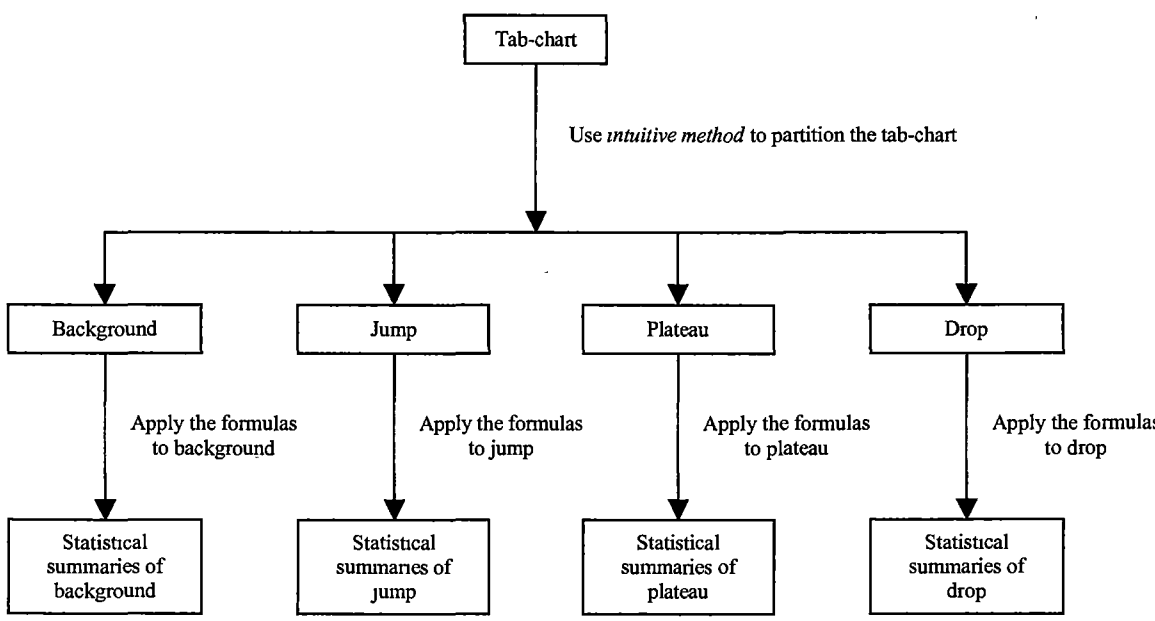


Fig 7.2 (1) : Overview And Summary Of Intuitive Method

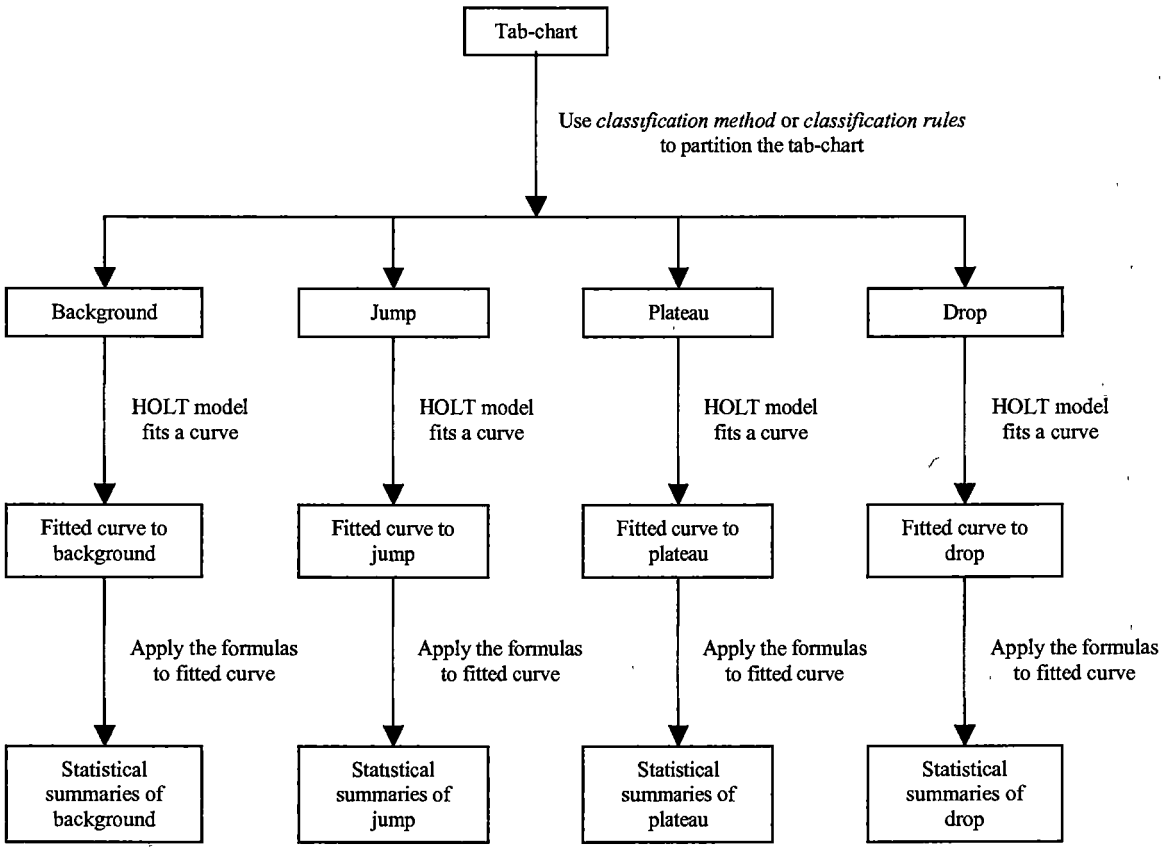


Fig 7.2 (2) : *Overview And Summary Of HOLT model, Classification Method And Rules*

(7.3) HOLT Model And Intuitive Method Comparison

(7.3.1) Standard Error Of Intuitive Method

The standard error of Intuitive method comes from the worksheet, which provided by the School of Earth Sciences at the University of Tasmania. The calculations of the worksheet are based on the content of their unpublished technical paper [9]. Although these standard errors possibly have other purposes, this section will use and borrow them to investigate the data reduction of HOLT model.

(7.3.2) Definition Of Standard Error Of HOLT Model And Intuitive Method

The quantitative analysis of data reduction requires definitions of standard errors of estimates. Two variants of standard error over background are used in the paper [9] : the first one is the standard error of background mean but the paper does not provide any derivation and explanation for second one. This second standard error seems to be the standard error of average amount of element. A standard error over plateau are used in the paper [9] : the standard error of plateau mean. The above standard errors are used in this project to compare the HOLT model and the intuitive method. These standard errors are listed in the following. The illustrative examples of calculating these standard errors are presented in the appendix (see Appendix F and G).

$$SE_1(BG) = \text{standard error of background mean} = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}}}$$

$$SE_2(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}}$$

$$SE(PL) = \text{standard error of plateau mean} = \sigma_{SG} \times \sqrt{\frac{1}{n_{SG}}}$$

For clear text and expression, when the standard error of background mean (i.e. $SE_1(BG)$) is calculated for the Intuitive method, the phrase *the standard error of Intuitive method over background* is used to represent the above meaning. When the standard error of background mean (i.e. $SE_1(BG)$) is calculated for the HOLT model, the phrase *the standard error of HOLT model over background* is used to represent the above meaning.

Similarly, when the standard error of plateau mean (i.e. SE(PL)) is calculated for the Intuitive method, the phrase *the standard error of Intuitive method over plateau* is used to represent the above meaning. When the standard error of plateau mean (i.e. SE(PL)) is calculated for the HOLT model, the phrase *the standard error of HOLT model over plateau* is used to represent the above meaning.

To avoid the confusion and mix-up, all statistical summaries (i.e. mean, standard deviation (error)) in this thesis are shown in the following table. In summary, the statistical summaries include the mean and standard deviation (error) of tab-chart, HOLT model and element concentration. Intuitive method, classification method and classification rules use the same formulas of statistical summaries but have different definitions in the notations of the formulas (see sections 4.3.5, 6.2.2, 6.2.3 and 6.3.1).

Fig 7.3 (1) : *Summary Of All Statistical Summaries In This Thesis. The first six statistical summaries (i.e. from (A) to (F)) are studied in the previous chapter, whereas the other statistical summaries (i.e. from (G) to (L)) are studied in this section*

Name Of Statistical Summary	*Notation*	*Calculation Formula*
(A) Mean (average level) of background	$\bar{x}_{BG}$	$\sum_{k\in\Phi} \frac{x_k}{n_{BG}}$
(B) Mean (average level) of plateau	$\bar{x}_{SG}$	$\sum_{k\in\Omega} \frac{x_k}{n_{SG}}$
(C) Standard deviation of background	σ_{BG}	$\sqrt{\frac{\sum_{k\in\Phi}(x_k-\bar{x}_{BG})^2}{n_{BG}-1}}$
(D) Standard deviation of plateau	σ_{SG}	$\sqrt{\frac{\sum_{k\in\Omega}(x_k-\bar{x}_{SG})^2}{n_{SG}-1}}$
(E) Mean (average level) of the element concentration	-----	$\lambda(\bar{x}_{SG}-\bar{x}_{BG})$
(F) Standard deviation of the element concentration	-----	$\lambda\sqrt{(\sigma_{BG}^2+\sigma_{SG}^2)}$
(G) Standard error of background mean	$SE_1(BG)$	$\sigma_{BG}\times\sqrt{\frac{1}{n_{BG}}}$
(H) Standard error of plateau mean	SE(PL)	$\sigma_{SG}\times\sqrt{\frac{1}{n_{SG}}}$
(I) Standard error of Intuitive method over background	-----	Use the observation values in the tab-chart over the integration interval for background to calculate $SE_1(BG)$
(J) Standard error of Intuitive method over plateau	-----	Use the observation values in the tab-chart over the integration interval for plateau to calculate SE(PL)
(K) Standard error of HOLT model over background	-----	Use the fitted values of HOLT model and the observations are classified as background by classification method to calculate $SE_1(BG)$
(L) Standard error of HOLT model over plateau	-----	Use the fitted values of HOLT model and the observations are classified as plateau by classification method to calculate SE(PL)

There is a point to be clarified and restated before moving to the next section. The HOLT model will use with classification method only in this chapter, the reason has given in the previous section (see section 7.2.2).

(7.3.3) ***The Problem Of The Standard Error Formulas***

Before starting the quantitative analysis of HOLT model's data reduction, there is a problem of the standard error formulas must be pointed out. The assumption of the formulas is that the raw data are independent. However, the raw data of some tab-charts are correlated because ARIMA can be fitted to these tab-charts (see section 8.4). So the formulas should be corrected to the following forms.

$$SE_1(BG) = \text{Standard error of background mean} = \sigma^*_{BG} \times \sqrt{\frac{1}{n_{BG}}}$$

$$SE_2(BG) = \sigma^*_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}}$$

where

$$\sigma^*_{BG} = \begin{cases} \sigma_{BG} & \text{if independent} \\ \dfrac{\sigma^2_\varepsilon}{1-\phi_1^2} & \text{if AR(1)} \end{cases}$$

and

$$SE(PL) = \text{Standard error of plateau mean} = \sigma^*_{SG} \times \sqrt{\frac{1}{n_{SG}}}$$

where

$$\sigma^*_{SG} = \begin{cases} \sigma_{BG} & \text{if independent} \\ \dfrac{\sigma^2_\varepsilon}{1-\phi_1^2} & \text{if AR(1)} \end{cases}$$

Note : If the tab-chart has other ARIMA structures, the σ^*_{BG} or σ^*_{SG} is replaced by the corresponding standard error.

(7.3.4) ***Data Reduction And Autocorrelation***

Although the standard error formulas from the technical paper [9] have problem when they are applied to the raw data (i.e. tab-chart) having the autocorrelation (see section 7.3.3), they are still used in studying the data reduction of HOLT model. The reasons are given at the following discussions. Firstly, the formula is not used to estimate the statistical summaries of the underlying process. It is used to measure the degree of variation or fluctuation in the background and plateau. Secondly, the data reduction of the HOLT model is capable of removing the variation and noise from the tab-chart. The researchers can gain the statistical information (i.e. summaries) of the tab-chart through the fitted curve (value) of the HOLT model. I agree that the HOLT model may give the biased estimation to the researchers due to the autocorrelation. If the autocorrelation is very serious, the tab-chart should be pre-processed to get rid of the autocorrelation before applying the HOLT model.

(7.3.5) ***Result Of HOLT Model And Intuitive Method Comparison***

For each element, the series plots (i.e. standard error of HOLT model and Intuitive method versus samples) over background and plateau are plotted. By visually comparing the standard error of HOLT model and Intuitive method over background, the standard error of Intuitive method is smaller than that of HOLT model. It is because signal intensity in the background of tab-chart is very close to zero or swap between zero and a set of close values. However, the situation is reversed when comparing them over plateau because the curve of HOLT model is smooth over the plateau but the tab-chart over the plateau is spiky.

The above phenomenon gives us the following inspiration. HOLT model loses the ability of data reduction over background and the HOLT model is unable to discriminate the background from the plateau of trace amount of element. In other words, it possibly gives us the false signal that the portion is plateau but it is actually a background. Therefore, HOLT model cannot replace the researcher's visual observation (i.e. Intuition method) in this background stage. Some methods should be explored to get rid of this problem existing in HOLT model. The first possible method is to use statistical test to find any significant deviation from zero level. The second possible method is to use a constant trend by choosing the appropriate parameters when the signal (i.e. counts per second) is very small (see section 5.3). In other words, the background (or plateau of small signal) and plateau should use different α and β parameters (see section 6.3.4).

The graphs of the standard error of HOLT model and Intuitive method in elements Nd146 and V51 are displayed in the following. The first and third graph (see Fig 7.3 (2) and Fig 7.3 (4)) are the series plot over background with the fitted value, $\hat{Y}_t(0) = L_t$ and the formula $SE_1(BG)$, whereas the second and fourth graph (see Fig 7.3 (3) and Fig 7.3 (5)) are the series plot over plateau with the fitted

value, $\hat{Y}_t(0) = L_t$. There are unusual high standard errors in the first few samples, around sample 30, around sample 40 and last few samples. These extraordinary values are possibly caused by the machine drift and the researchers use the primary standards to correct and calibrate the machine.

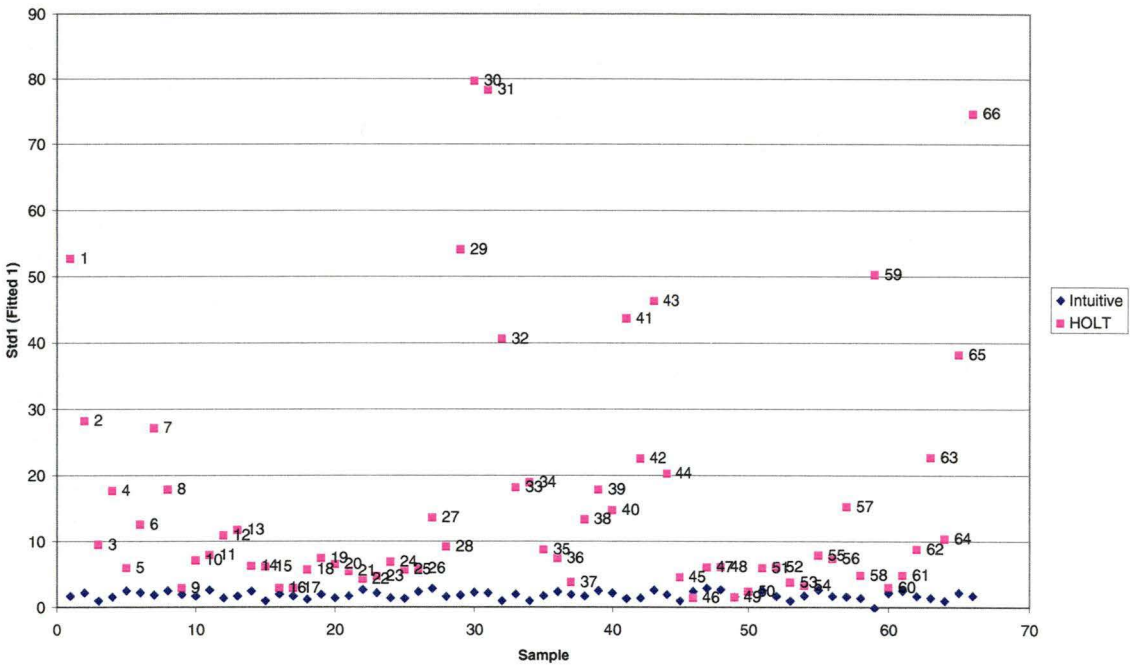


Fig 7.3 (2) : *Series Plot Of HOLT Versus And Intuitive Method Over Background (Nd146). The blue diamond represents intuitive method, whereas the pink rectangle represents HOLT model with* $\hat{Y}_t(0)$

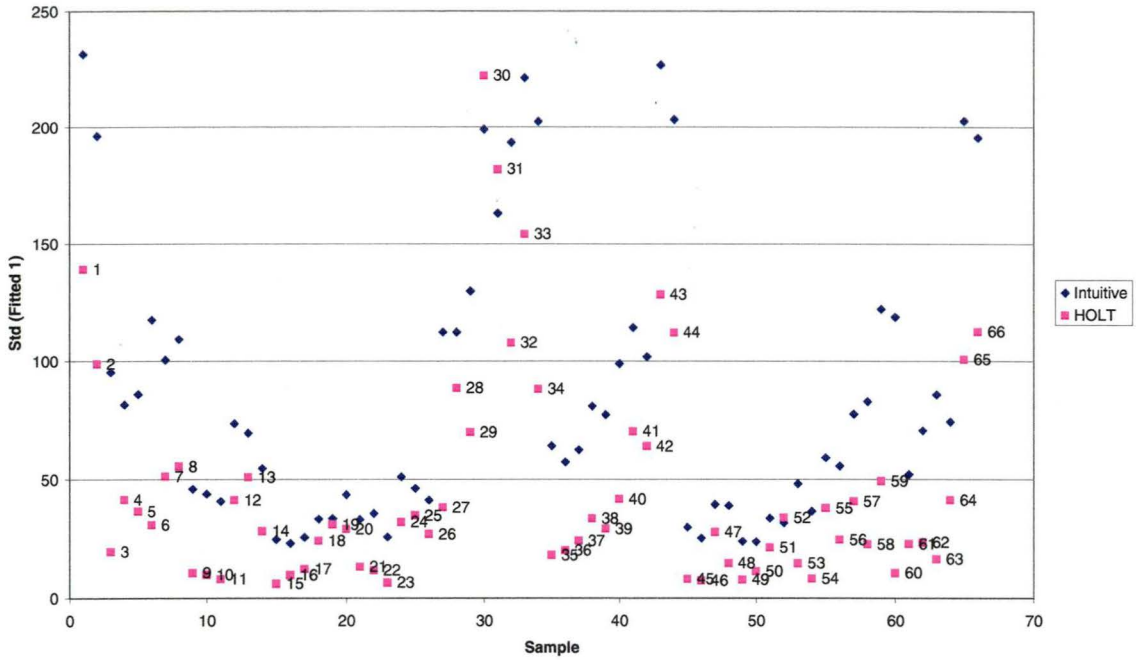


Fig 7.3 (3) : *Series Plot Of HOLT And Intuitive Method Over Plateau (Nd146). The blue diamond represents intuitive method, whereas the pink rectangle represents HOLT model with* $\hat{Y}_t(0)$

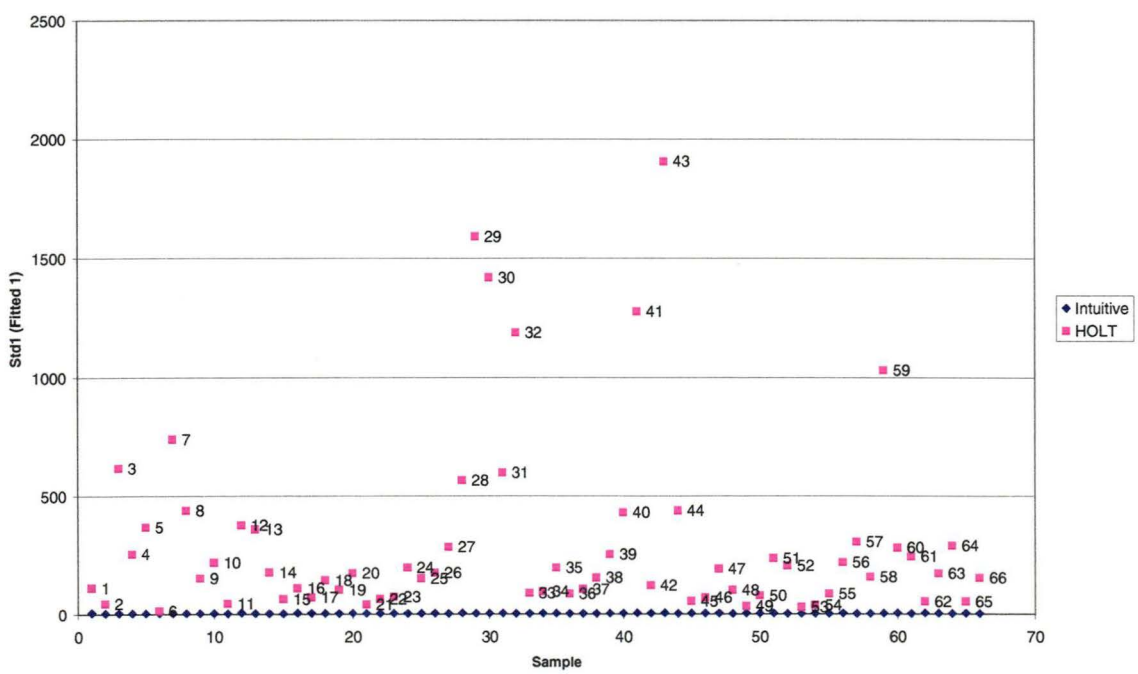


Fig 7.3 (4) : *Series Plot Of HOLT Versus And Intuitive Method Over Background (V51). The blue diamond represents intuitive method, whereas the pink rectangle represents HOLT model with* $\hat{Y}_t(0)$

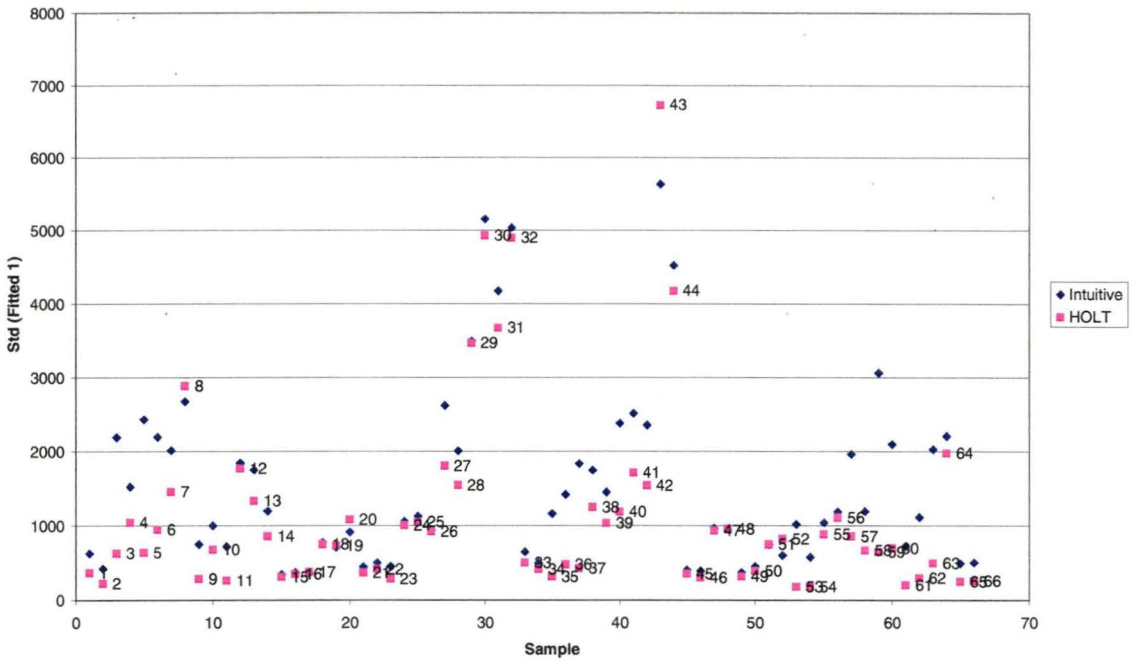


Fig 7.3 (5) : *Series Plot Of HOLT Versus And Intuitive Method Over Plateau (V51). The blue diamond represents intuitive method, whereas the pink rectangle represents HOLT model with* $\hat{Y}_t(0)$

Then the above series plots are changed into scatterplots in order to check whether there is some relationships between noise variations of Intuitive method and HOLT model. After studying the scatterplots of all elements (i.e. total 28 elements), the conclusions are drawn at the following. Over the background, the first finding is that the scale of y-axis is larger than that of x-axis in most of the scatterplots. The second finding is that the regression lines in most of the scatterplots have negative slope or great positive slope. These quantitative findings support that the standard error of Intuitive method is smaller than that of HOLT model (see Fig 7.3 (6) and Fig 7.3 (8)). Over the plateau, the slope of regression lines in most of the scatterplots is positive and less than one. It means that the standard error of Intuitive method is greater than that of HOLT model (see Fig 7.3 (7) and Fig 7.3 (9)). But there is a greater range of values for both methods, and they do increase together (i.e. there is a relationship, or they are in reasonable agreement).

There is a remark to explain the regression line in the graph. Since the tab-charts violate the independent and equal variance conditions in the linear regression, there are several serious consequences of using the ordinary least squares procedure, including confidence interval, hypothesis tests and predictions, etc. However, the estimated regression coefficients in the regression are still unbiased, but they have not the minimum variance property and may be quite inefficient [3], [17]. Besides, the regression line in the graph is used as a tool to support and clarify the idea of data reduction of HOLT model only.

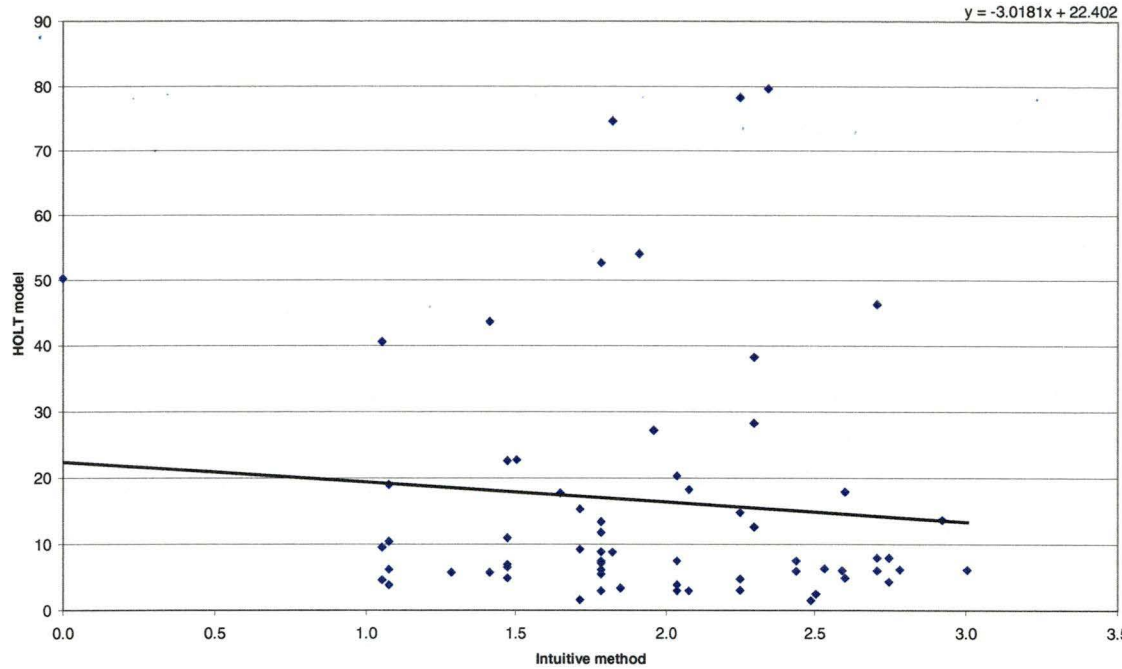


Fig 7.3 (6) : *Scatterplot Of HOLT Model Versus Intuitive Method Over Background (Nd146). The points are* $SE_1(BG)$ and *the solid line is regression line :* $y = -3.0181\,x + 22.402$

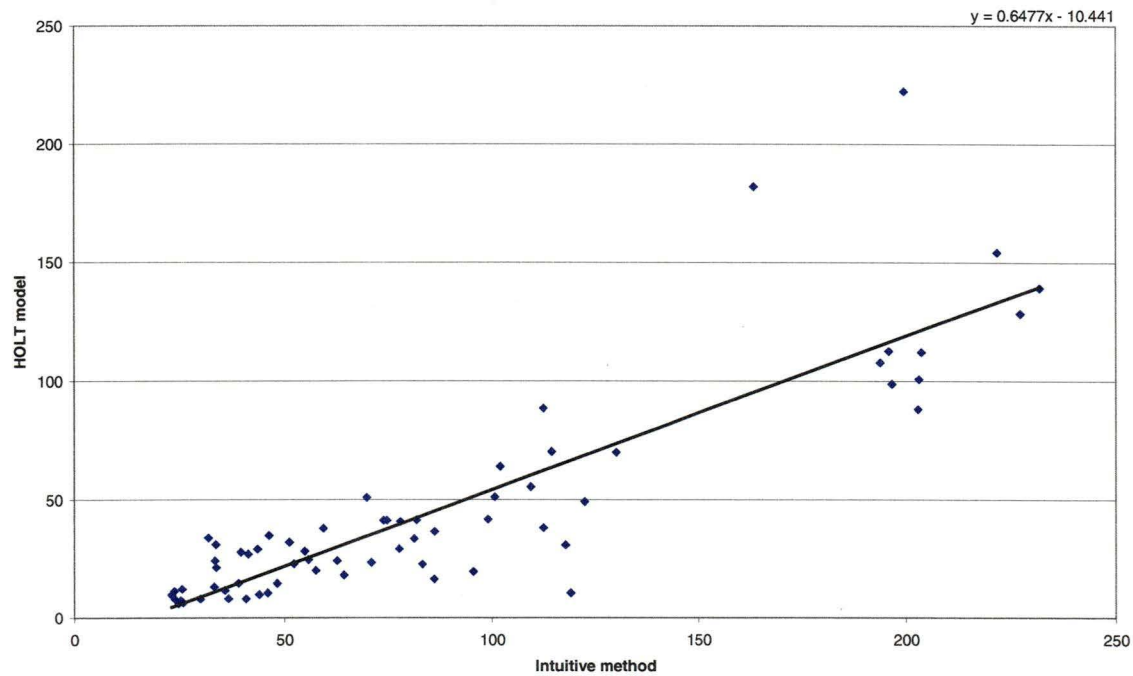


Fig 7.3 (7) : *Scatterplot Of HOLT Model Versus Intuitive Method Over Plateau (Nd146). The point are* $SE(PL)$ *and the solid line is regression line :* $y = 0.6477\,x - 10.441$

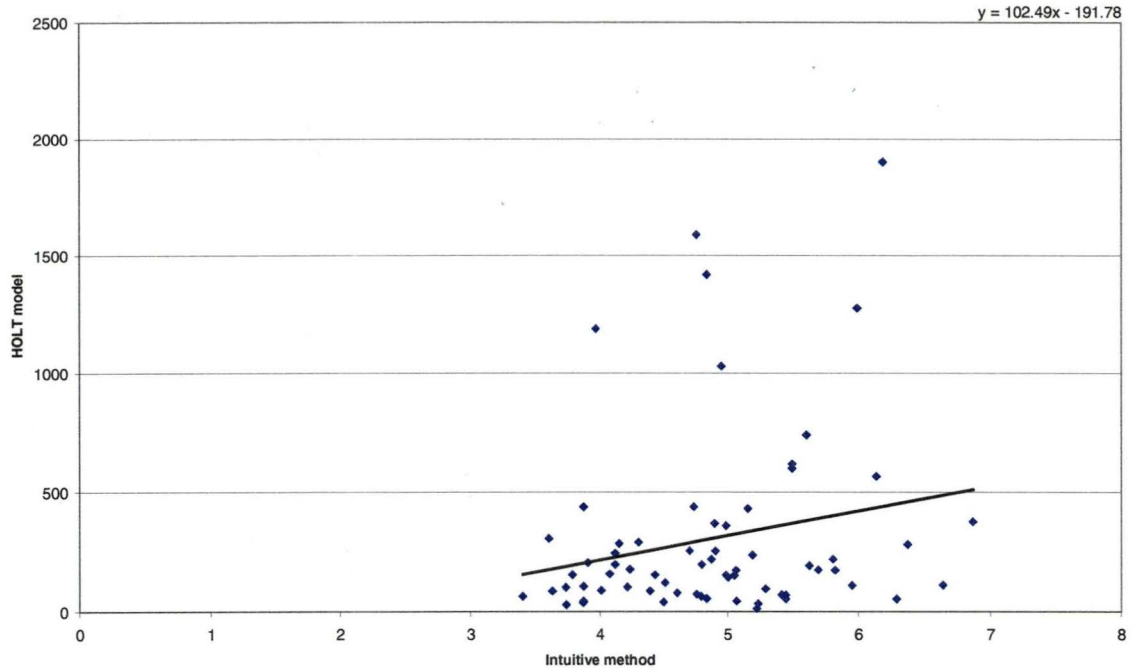


Fig 7.3 (8) : *Scatterplot Of HOLT Model Versus Intuitive Method Over Background (V51). The points are* $SE_1(BG)$ and *the solid line is regression line :* $y = 102.49\,x - 191.78$

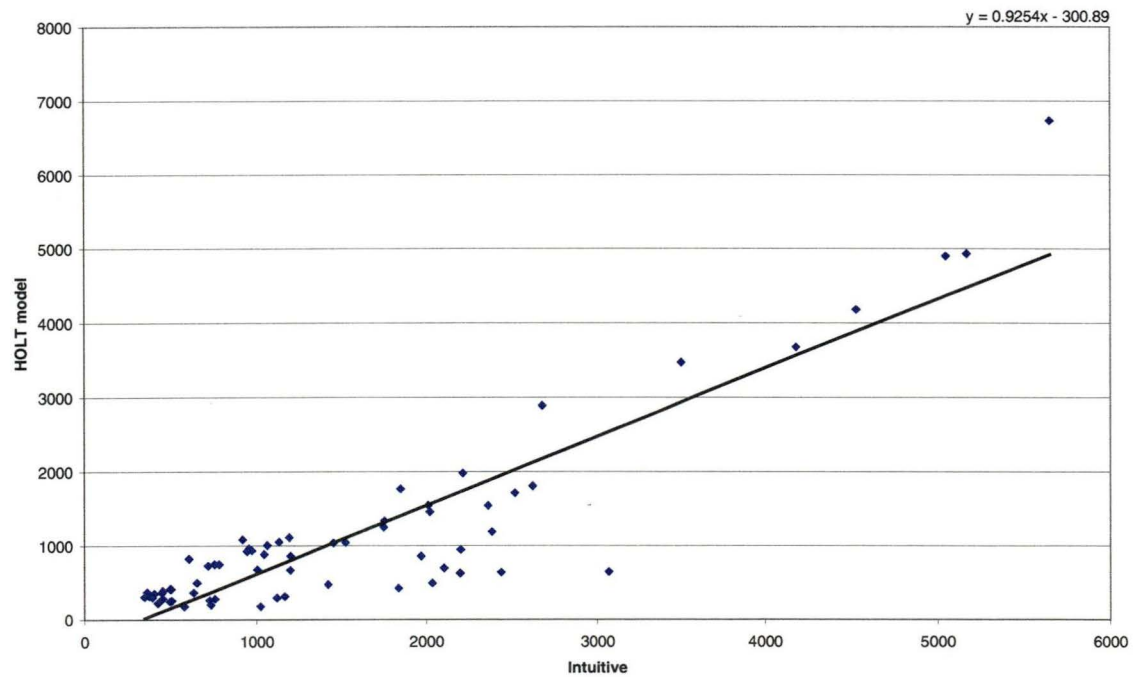


Fig 7.3 (9) : *Scatterplot Of HOLT Model Versus Intuitive Method Over Plateau (V51). The point are* $SE(PL)$ *and the solid line is regression line :* $y = 0.9254\,x - 300.89$

(7.4) Statistical Analysis Of HOLT Model

(7.4.1) Standard Deviation Of Fitted Error

When fitting the time series model to the raw data, the analysis of the fitting error should also be provided. The analysis can give the researchers the picture of the fitting error in different situations. Therefore, they can adjust the parameters to meet their requirement (i.e. decrease (increase) the fitting error but increase (decrease) the data reduction of HOLT model) before applying it to the tab-chart. Another purpose of studying the standard deviation of fitting error will be given later in this section (see section 7.4.6).

(7.4.2) Definition Of Standard Deviation Of Fitting Error

The standard deviation of fitting error is used to measure the performance of HOLT model (i.e. degree of fitness to tab-chart). First of all, each (actual) observation value is subtracted by fitted value (see section 5.4.2) from the classification method. Then we evaluate the standard deviation of these differences. The equations are given below. Some notations are already defined in previous sections (see sections 4.3.5, 5.4.2 and 6.2.2).

$$\text{Fitting error : } \begin{cases} e_t^1 = Y_t - \hat{Y}_t(0) = Y_t - L_t \\ \\ e_t^2 = Y_t - \hat{Y}_{t-1}(1) = Y_t - (L_{t-1} + T_{t-1}) \end{cases}$$

$$\text{Standard deviation of fitting error : } \begin{cases} Std_1(BG) = \sqrt{\dfrac{1}{n_{BG}-1}\sum_{t\in\Phi}(e_t^1 - \overline{e_t^1})^2} \\ \\ Std_1(SG) = \sqrt{\dfrac{1}{n_{SG}-1}\sum_{t\in\Omega}(e_t^1 - \overline{e_t^1})^2} \\ \\ Std_2(BG) = \sqrt{\dfrac{1}{n_{BG}-1}\sum_{t\in\Phi}(e_t^2 - \overline{e_t^2})^2} \\ \\ Std_2(SG) = \sqrt{\dfrac{1}{n_{SG}-1}\sum_{t\in\Omega}(e_t^2 - \overline{e_t^2})^2} \end{cases}$$

where

Y_t = the count per second (i.e. signal intensity or observation value) of the tab-chart

$Std_1(BG)$ = the standard deviation of fitting error over background using $\hat{Y}_t(0)$

$Std_1(SG)$ = the standard deviation of fitting error over plateau using $\hat{Y}_t(0)$

$Std_2(BG)$ = the standard deviation of fitting error over background using $\hat{Y}_{t-1}(1)$

$Std_2(SG)$ = the standard deviation of fitting error over plateau using $\hat{Y}_{t-1}(1)$

There are two remarks. The first remark is that $Std_1(SG)$ and $Std_2(SG)$ are studied only in this section because most of the observations over the background are zero or very small. However, it is reasonable and justified that the analysis result of $Std_1(SG)$ and $Std_2(SG)$ is also true in $Std_1(BG)$ and $Std_2(BG)$.

The second remark is that the standard deviation in this section is totally different from the standard deviation and error in the previous section (see section 7.3). In the previous section, the calculation of the standard deviation and error (i.e. σ_X and SE(X)) is based on the level of the tab-chart or observation value. In this section, the calculation of the standard deviation (i.e. Std(X)) is based on the difference of observation value and fitted value (see Fig 7.3 (1) and Appendix J).

(7.4.3) Relationship Between Standard Deviation Of Fitting Error And Average Level Of Plateau

The (approximate) average level of plateau is estimated by averaging the fitted value (i.e. $\hat{Y}_{t-1}(1)$ or $\hat{Y}_t(0)$), over the portion, which is classified as plateau by HOLT model together with the classification method. If the HOLT model fits well to the tab-chart, the fitting and classification is judged to be satisfactory. The equations are already shown in previous section (see sections 6.2.2).

After processing all samples and elements, the graphs of standard deviation of fitting error against average level of plateau are plotted for all elements. We found that standard deviation of fitting error increases as average level of plateau increases in all elements. For example, the following charts are the plots of standard deviation of fitting error versus average level of plateau in element Sc45. Other elements present a similar picture. The up-hill slope of regression lines supports the above argument.

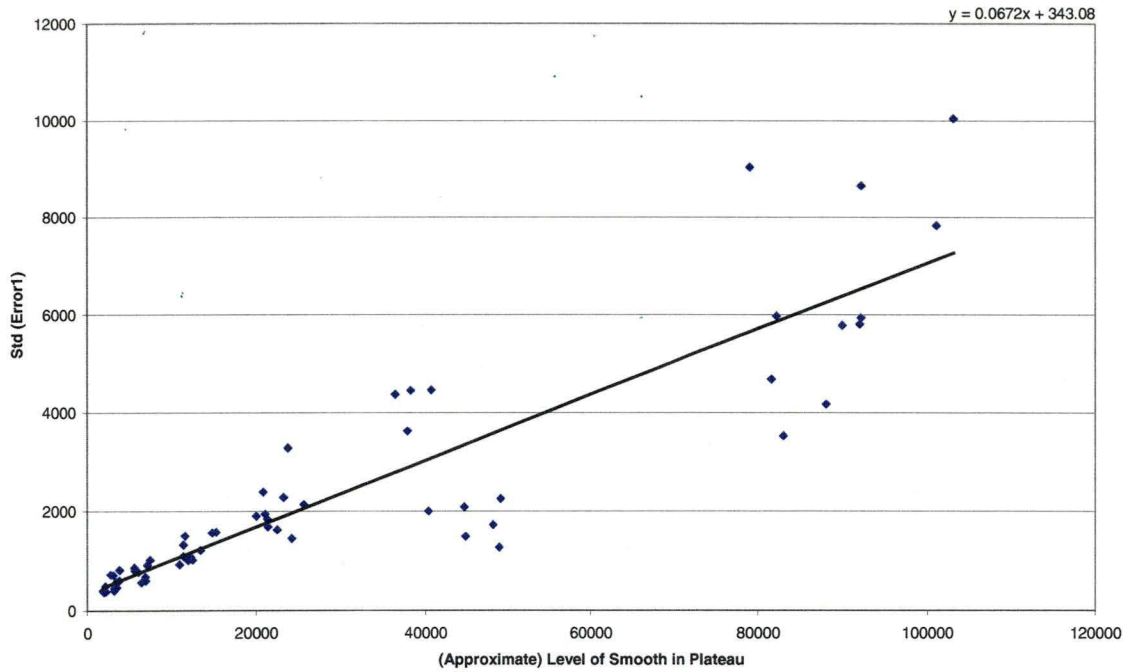


Fig 7.4 (1) : *Plot Of $Std_1(SG)$ Versus Average Level Of Plateau (Sc45). The solid line is regression line : $y = 0.0672\,x + 343.08$*

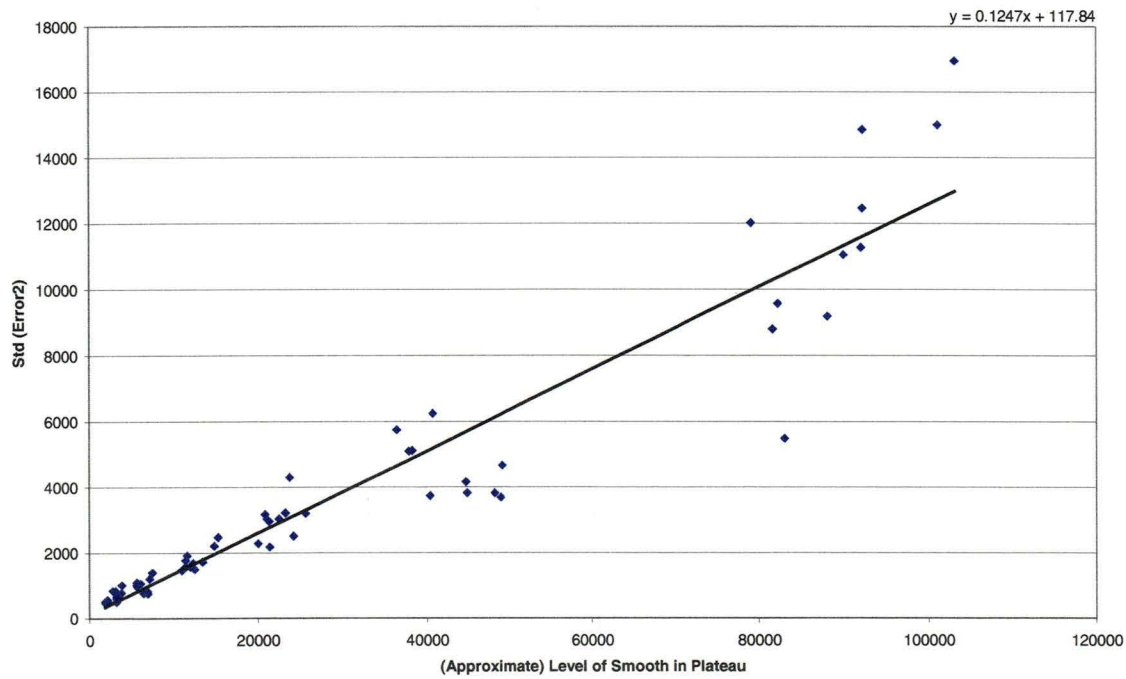


Fig 7.4 (2) : *Plot Of $Std_2(SG)$ Versus Average Level Of Plateau (Sc45). The solid line is regression line : $y = 0.1247\,x + 117.84$*

(7.4.4) Relationship Between Standard Deviation Of Fitting Error And Mass Number

The graphs of standard deviation of fitting error against mass number are plotted for all samples. We found that the standard deviation of fitting error decreases as the mass number increases in all samples (i.e. decreases from the light element to heavy element). It implies that the HOLT model has better performance in heavy element than light element. It is because the amount of light element is higher than that of heavy element (see section 6.2.2), so the signal and average level of plateau of light element is higher than that of heavy element (see section 7.4.3). Therefore, the standard deviation of fitting error actually depends on the average level of plateau only. The following charts are the plots of standard deviation of fitting error versus mass number in sample 11.

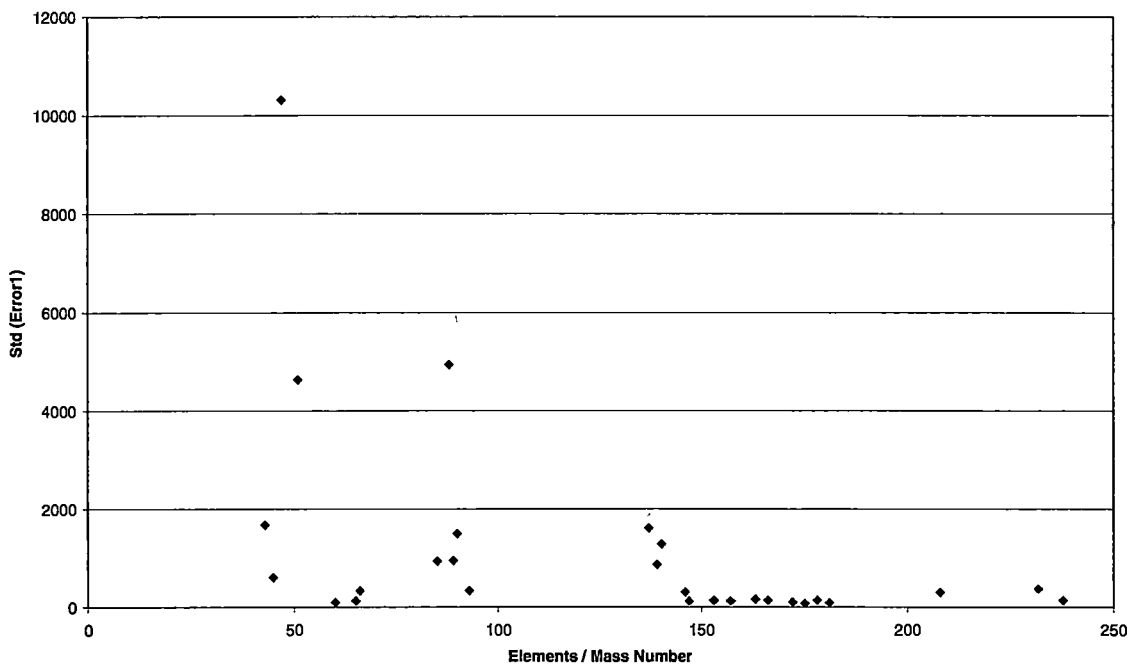


Fig 7.4 (3) : Plot Of $Std_1(SG)$ Versus Mass Number (Sample 11)

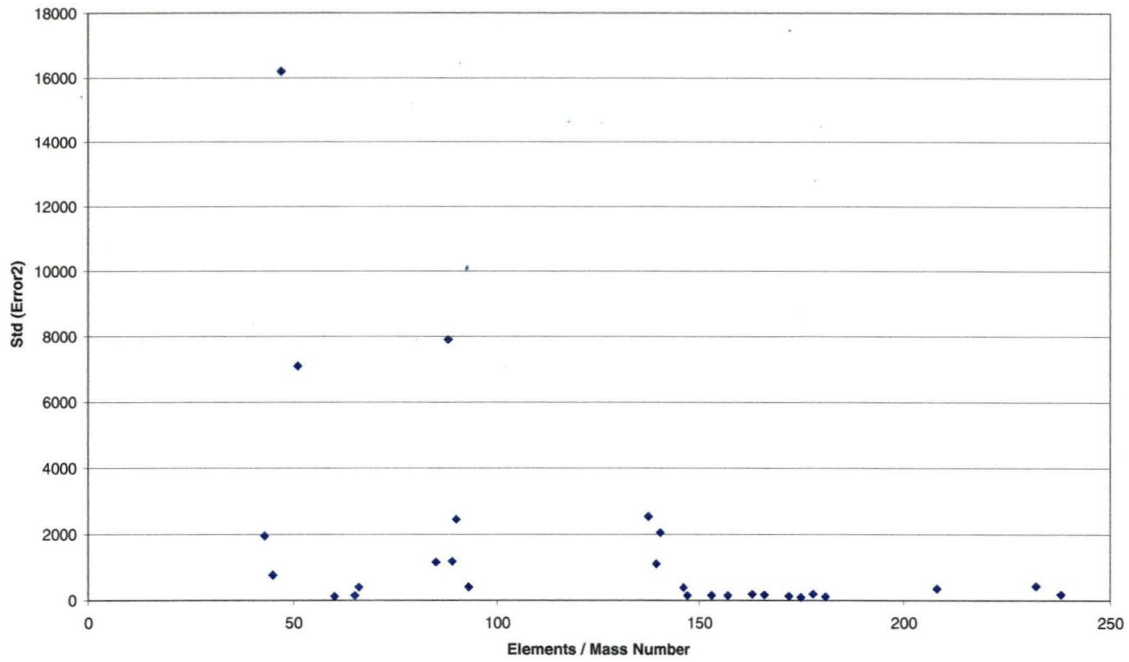


Fig 7.4 (4) : *Plot Of* $Std_2(SG)$ *Versus Mass Number (Sample 11)*

(7.4.5) ***Regression : Standard Deviation Of Fitting Error, Average Level Of Plateau & Mass Number***

As mentioned in previous section (see section 7.3.5), the tab-charts violate the independent and equal variance conditions in the linear regression, there are several serious consequences of using the ordinary least squares formula. However, the estimated regression coefficients in the regression are still unbiased and reliable. So the analysis in this section studies the coefficients only but ignore the other statistical values (eg. standard error, t statistics, P-value and ANOVA tables) (see section 7.3.5) [3], [17].

This section is different from the previous sections 7.4.3 and 7.4.4. The data of all elements are pooled for the regression. By regressing standard deviation of fitting error on average level of plateau, mass number and the interaction (i.e. average level of plateau × mass number), we obtain the regression equations in the following. The term Avg Level of Plateau has positive coefficient. These coefficients consistent with the previous results, which standard deviation of fitting error increases with average level of plateau. Since other elements have the similar results after studying, one regression line for all elements can avoid a lot of graphs of each element shown in this thesis.

$$\begin{cases} \text{Std}_1(\text{SG}) = 1650 + 0.057566 \times \text{Avg Level of Plateau} - 7.5861 \times \text{Mass No} \\ \qquad\qquad + 0.00007102 \times \text{Avg Level of Plateau} \times \text{Mass No} \\ \text{Std}_2(\text{SG}) = 1700 + 0.097545 \times \text{Avg Level of Plateau} - 7.6330 \times \text{Mass No} \\ \qquad\qquad + 0.00014870 \times \text{Avg Level of Plateau} \times \text{Mass No} \end{cases}$$

Fig 7.4 (5) : *Table Of Regressing $Std_1(SG)$ On Plat Level, Mass No And Interaction. Note that the values in standard error, t Stat and P-value columns are not reliable and all values are corrected to six significant figures*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1650.14	264.293	6.24361	$< 10^{-9}$
Level of Plat	0.057566	0.001375	41.8575	0
Mass No	-7.58610	1.87321	-4.04978	$< 10^{-4}$
Level of Plat × Mass No	$< 10^{-4}$	$< 10^{-4}$	3.19834	0.001406

Fig 7.4 (6) : *Table Of Regressing $Std_2(SG)$ On Plat Level, Mass No And Interaction. Note that the values in standard error, t Stat and P-value columns are not reliable and all values are corrected to six significant figures*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1700.14	266.748	6.37359	$< 10^{-9}$
Level of Plat	0.097545	0.001388	70.2744	0
Mass No	-7.63297	1.89061	-4.03730	$< 10^{-4}$
Level of Plat × Mass No	0.000149	$< 10^{-4}$	6.63436	$< 10^{-10}$

The interaction term (i.e. Avg Plateau Level × Mass No) can be neglected because the term *0.00007102 × Mass No* in the regression equation $\text{Std}_1(\text{SG})$ and *0.00014870 × Mass No* in the regression line $\text{Std}_2(\text{SG})$ have maximum values *0.00007102 × 238 = 0.01690* and *0.00014870 × 238 = 0.03539*. These values (i.e. 0.01690 and 0.03539) are relatively small compared to the coefficients of Avg Level of Plateau (i.e. 0.057566 and 0.097545), especially when the term Avg Level of Plateau is large. Thus the term Avg Level of Plateau will dominate the interaction term in the regression equation. By the same argument, if the term Avg Level of Plateau is sufficiently large, it will also dominate the term Mass No.

So for practical purposes, the regression analysis is redone by (a) regressing standard deviation of fitting error on average level of plateau and mass number (see Fig 7.4 (7) and Fig 7.4 (8)) and (b) regressing standard deviation of fitting error on average level of plateau only (see Fig 7.4 (9) and Fig 7.4 (10)). The ANOVA tables are list as the following.

Fig 7.4 (7) : *Table Of Regressing* $Std_1(SG)$ *On Plat Level And Mass No. Note that the values in standard error, t Stat and P-value columns are not reliable and all values are corrected to six significant figures*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1594.08	264.370	6.02972	$< 10^{-8}$
Level of Plat	0.061576	0.000567	108.693	0
Mass No	-6.39694	1.84053	-3.47560	0.000521

Fig 7.4 (8) : *Table Of Regressing* $Std_2(SG)$ *On Plat Level And Mass No. Note that the values in standard error, t Stat and P-value columns are not reliable and all values are corrected to six significant figures*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1582.77	269.246	5.87851	$< 10^{-8}$
Level of Plat	0.105940	0.000577	183.618	0
Mass No	-5.14338	1.87447	-2.74391	0.006130

Fig 7.4 (9) : *Table Of Regressing* $Std_1(SG)$ *On Plat Level. Note that the values in standard error, t Stat and P-value columns are not reliable and all values are corrected to six significant figures*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	752.163	106.211	7.08179	$< 10^{-11}$
Level of Plat	0.062092	0.000548	113.236	0

Fig 7.4 (10) : *Table Of Regressing* $Std_2(SG)$ *On Plat Level. Note that the values in standard error, t Stat and P-value columns are not reliable and all values are corrected to six significant figures*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	905.835	108.037	8.38448	$< 10^{-16}$
Level of Plat	0.106355	0.000558	190.679	0

From the above two ANOVA tables (see Fig 7.4 (9) and Fig 7.4 (10)), the term standard deviation of fitting error is approximately proportional to average level of plateau. By inspection, no curve-linear fit is better than a linear one. The equations of the model are

$$\begin{cases} \text{Std}_1(\text{SG}) = 752.16 + 0.062 \times \text{Avg Level of Plateau} \\ \text{Std}_2(\text{SG}) = 905.84 + 0.106 \times \text{Avg Level of Plateau} \end{cases} \qquad (\Phi 1)$$

The mass spectrometer output is proportional to the number of nuclei, which one might expect to be approximately Poisson distributed in a discrete sample approximation. The above analysis however contradicts the $\text{SD} \propto \sqrt{\text{Mean}}$ behaviour of Poisson count; hence the spectrometer errors must have other sources (rather than just Poisson variation). Since it is reasonable that the result of plateau is also true in background, the above equations (Φ1) can be rewritten into the following general form (see the first remark in section 7.4.2).

$$\begin{cases} \text{Std}_1(\text{Avg Level}) = 752.16 + 0.062 \times \text{Avg Level} \\ \text{Std}_2(\text{Avg Level}) = 905.84 + 0.106 \times \text{Avg Level} \end{cases} \qquad (\Phi 2)$$

(7.4.6) Standard Deviation Of Fitting Error Of The Element Concentration

After studying the fitting error of plateau, the study will turn to the standard deviation of fitting error of element concentration. The equation is given at below. There is an assumption that the background and plateau are independent. For simple notations, it also assumes that the actual amount of element is equal to the concentration of element (i.e. set $\lambda = 1$) (see section 4.3.3).

$$\begin{aligned} \text{Std}_m(\text{Concentration}) &= \sqrt{\text{Std}_m^2(\text{cps}_{SG} - \text{cps}_{BG})} \\ &= \sqrt{\text{Std}_m^2(\text{SG}) + \text{Std}_m^2(\text{BG})} \end{aligned} \qquad (\Psi)$$

where

(a) $\text{Std}_m^2(\text{BG})$ = square value of $\text{Std}_m(\text{BG})$

(b) $\text{Std}_m^2(\text{SG})$ = square value of $\text{Std}_m(\text{SG})$

(c) $m = 1$ (i.e. $\hat{Y}_t(0)$) or 2 (i.e. $\hat{Y}_{t-1}(1)$) (from classification method)

In the equation (Ψ), if the signal or count per second of the background is sufficiently small, we have $Std_m(\text{Concentration}) \approx Std_m(SG)$ because we have standard deviation of fitting error is approximately proportional to average level in the previous section (see equation (Φ2) in section 7.4.5).

The result of the last paragraph has the following implications. If the researcher or machine manufacturer can minimizes the effect of the background noise, which depends on the machine and can be controlled, then $Std_m(BG)$ (i.e. $Std_m^2(BG)$) is minimized by equation (Φ2) in the previous section. So the standard deviation of fitting error of element concentration (i.e. $Std_m(\text{Concentration})$) can be calculated by considering the term $Std_m(SG)$ only (i.e. ignoring the term $Std_m(BG)$). Therefore, the confidence interval of the average element concentration will be more accurate.

However, when the signal intensity of plateau is much greater than background noise (i.e. Avg Level of background << Avg Level of plateau), the control of the background noise is no longer helpful because $Std_m(\text{Concentration}) \approx Std_m(SG)$. To sum up, the control of background noise can help to estimate the trace of the element in the sample more accurate.

(7.5) Data Reduction, Fitting Error, Parameters And Performance Of HOLT

After the studies and investigation in this and previous chapters, there is an interesting point to be raised. Obviously, the data reduction and fitting error of the HOLT model counterbalance or offset each other. By changing the parameter values and threshold value, the data reduction will increase (decrease) but the fitting error will decrease (increase). In other words, the researchers can gain (lose) information of the trend of the tab-chart but obtain the well-fitted (worse-fitted) curve to the tab-chart.

On the other hand, the parameter values may affect the performance of the classification method and rules to some extent, even although this project does not carry out experiment to verify this fact and there are other factors (i.e. the algorithm and the set of rules) affecting the performance. There is a complicated relationship between data reduction, fitting error, choice of parameters and performance.

(7.6) Chapter Conclusions

The first part of this chapter is to study the data reduction of HOLT model. The standard error of observation (fitted value) mean can be used to quantify the smoothness (or fluctuation) of the Intuitive method (HOLT model). The *square-root of the sample size formula* is still employed to calculate the standard error of the background and plateau although most of the tab-charts are subject to autocorrelation. The reason is that the formula is not used to estimate the statistical summaries of the underlying process; it is used to measure the degree of variation or fluctuation in the background and plateau. From the result of the analysis, the HOLT model can trace the trend of the tab-chart and is capable of removing the fluctuation from the tab-chart (i.e. data reduction). However, the data reduction of the HOLT model does not work on the background or the plateau of small signal or count per second (i.e. trace amount of element in the sample). This drawback is a hint to how to improve the data reduction of the HOLT model. The possible solution is that different parameter values are applied to background and plateau since our current policy is that the same parameter values are for both background and plateau.

The second part of this chapter is to study the fitting error of HOLT model to the tab-chart. This part focuses on the analysis of the standard deviation of fitting error over the plateau only. The standard deviation of fitting error over the background is not considered because most of the observations over the background are zero or very small. However it is believable that the conclusion of plateau should also be true for background. By the analysis result, the standard deviation of fitting error over the plateau is proportional to the average level of the plateau. This finding implies that minimizing the background (noise) will make the confidence interval of the average concentration of element more accurate. However, it the plateau (signal) is much higher than the background (noise), the control of the background (noise) is no longer helpful; it is capable of delivering only minor improvement. The control of background noise can help to estimate the trace of the element in the sample more accurate.

The data reduction and fitting error of the HOLT model counterbalance (i.e. offset) each other and are controlled by the parameters. In other words, the researchers can adjust the parameters to decrease (increase) the fitting error but increase (decrease) the data reduction of the HOLT model before applying the model to the tab-chart. In the further studies, an experiment should be carried out to investigate the relationship between the parameters and fitting error of the model.

Chapter 8 ARIMA Modelling Of The Tab-Chart

(8.1) The Work Of This Chapter

For the comprehensive analysis, the ARIMA model could also be applied to the tab-chart after the time series model is already applied in the previous chapters. This chapter devises procedures to speed up the ARIMA model fitting because there are a lot of tab-charts (total 1848 tab-charts) to be fitted. Then the ARIMA model is fitted to the background and plateau of the tab-chart. The last work of this chapter is to discuss the suitability of ARIMA model in this project and compare the ARIMA model and HOLT model.

(8.2) Introduction To ARIMA Model

Apart from exponential smoothing method, ARIMA model is an alternative and famous method in data fitting and forecasting. Box and Jenkins proposed this model in 1967. Both methods have similar features. They are extrapolative and adaptive methods and can model the seasonality. Actually, exponential smoothing method is equivalent to ARIMA(0,1,1) model without constant term, see [27].

However, there are two differences between these two methods. Firstly, ARIMA requires the examination of the autocorrelation function (ACF) and partial autocorrelation (PACF). Secondly, exponential smoothing method has both linear and nonlinear models whereas ARIMA has linear model only. For more detail of various ARIMA models, see [1].

(8.3) Procedures Of Fitting ARIMA To Tab-Chart

In this project, the ARIMA model is fitted to the background and plateau portion of the tab-chart. For each sample, the series is checked for stationary and the autocorrelation function (ACF) and partial autocorrelation (PACF) are plotted in the first step to identify appropriate ARIMA models. If the plots show non-stationary, differencing will be carried out on the tab-chart. Since there are over a thousand tab-charts to be fitted by ARMA model, a systematic method is used in order to speed up the process. Five ARMA models (i.e. AR(1), AR(2), MA(1), MA(2) and ARMA(1,1)) are tried to fit to the plateau of the tab-charts, the best model is chosen by the lowest MSE and significant parameters.

(8.4) Findings And Results

(8.4.1) Background Portion Of The Tab-Chart

After the fitting process, the following results are obtained. No differencing is needed to apply on all tab-charts. The largest number of samples are ARMA(1,1) and the second largest number of samples are random process or inconclusive because the five models cannot be fitted to these tab-charts. Two typical analyses are shown below, in Fig 8.4 (1) to Fig 8.4 (4). The first example at below is element La139 and sample 43. The best model for this sample is ARMA(1,1). The second example is Lu175 and sample 39. The best model for this sample is also ARMA(1,1)

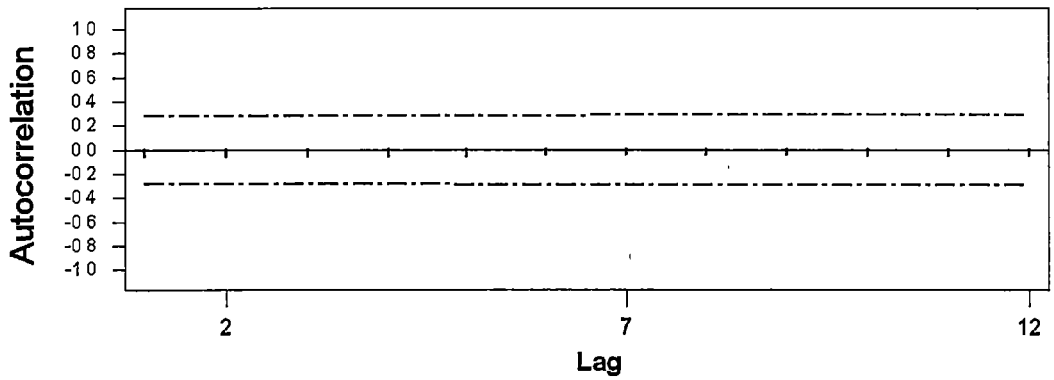


Fig 8.4 (1) : ACF Plot Of La139, Sample 43 (Background). There is no significance spike in ACF

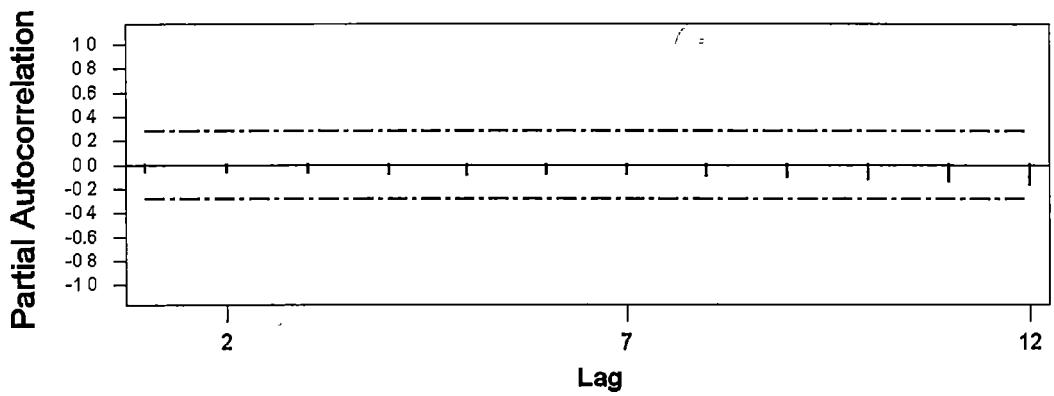


Fig 8.4 (2) : PACF Plot Of La139, Sample 43 (Background). There is no significance spike in PACF

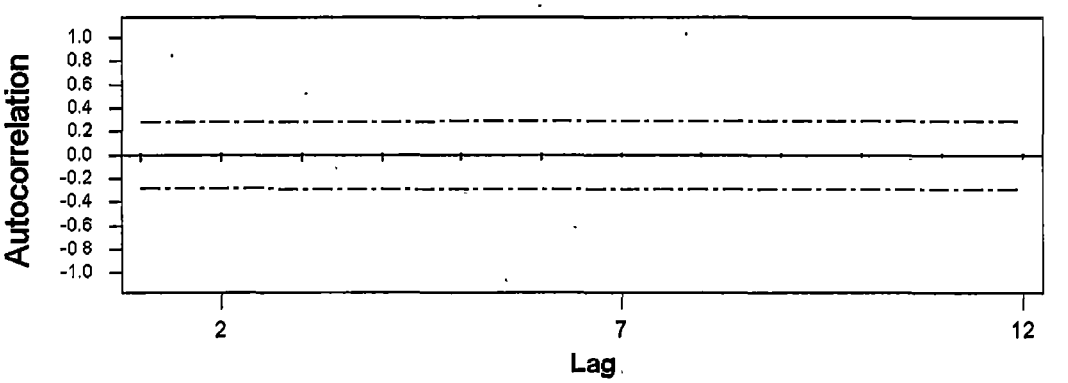


Fig 8.4 (3) : *ACF Plot Of Lu175, Sample 39 (Background). There is no significance spike in ACF*

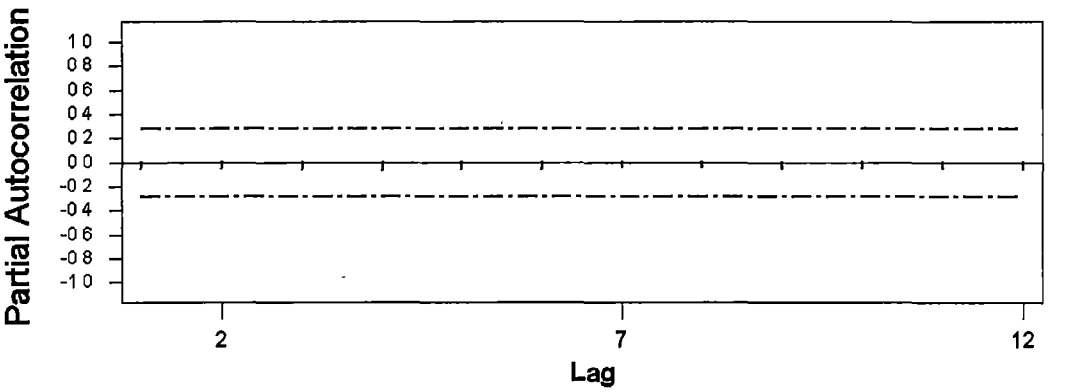


Fig 8.4 (4) : *PACF Plot Of Lu175, Sample 39 (Background). There is no significance spike in PACF*

(8.4.2) ***Plateau Portion Of The Tab-Chart***

After the fitting process, the following results are obtained. The first difference is adequate for all non-stationary tab-charts. For stationary sample, most of the tab-charts are AR(1) or ARMA(1,1). For the non-stationary tab-charts, most of the tab-charts are ARIMA(0,1,1). The rest of the tab-charts are random process or inconclusive because the five models cannot be fitted to these tab-charts. Two typical analyses are shown below, in Fig 8.4 (5) to Fig 8.4 (12). The first example at below is element La139 and sample 43. The best model for this sample is ARIMA(0,1,1). The ACF is non-stationary since there is an upward trend in the tab-chart (see Fig 8.4 (5) to Fig 8.4 (9)). The second example is Lu175 and sample 39. The best model is AR(1) and there is a horizontal trend in the tab-chart (see Fig 8.4 (10) to Fig 8.4 (12)).

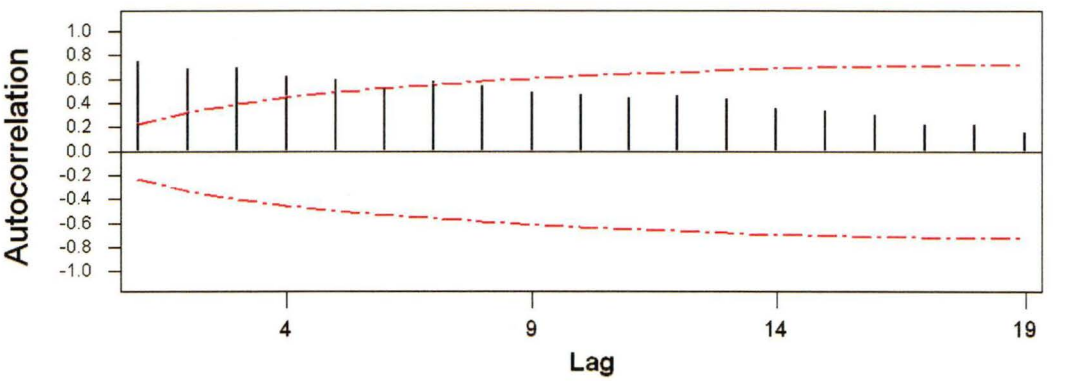


Fig 8.4 (5) : *ACF Plot Of La139, Sample 43 (Plateau). The spikes in ACF decrease very slowly*

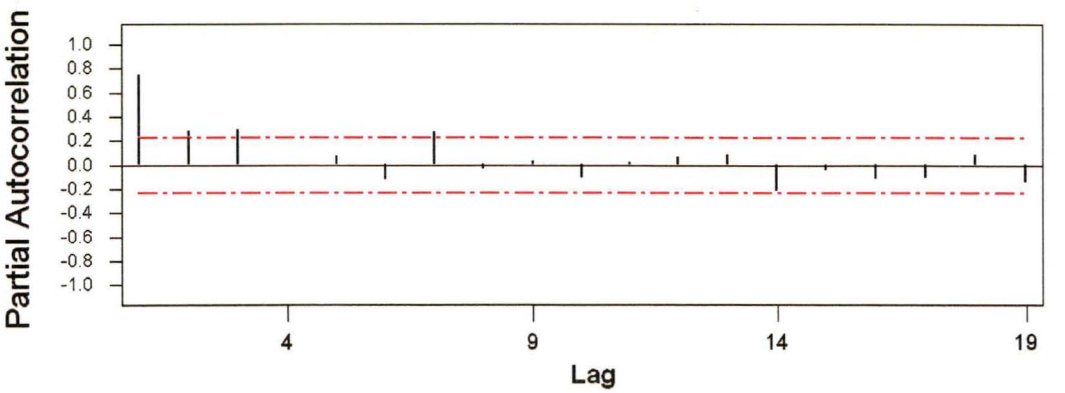


Fig 8.4 (6) : *PACF Plot Of La139, Sample 43 (Plateau). There is a cut off after the third spike*

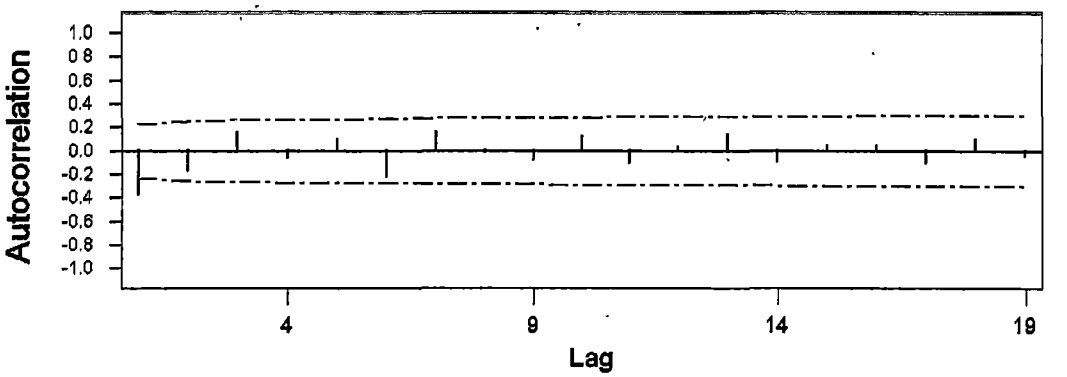


Fig 8.4 (7) : ACF Plot Of La139 After First Difference, Sample 43 (Plateau).

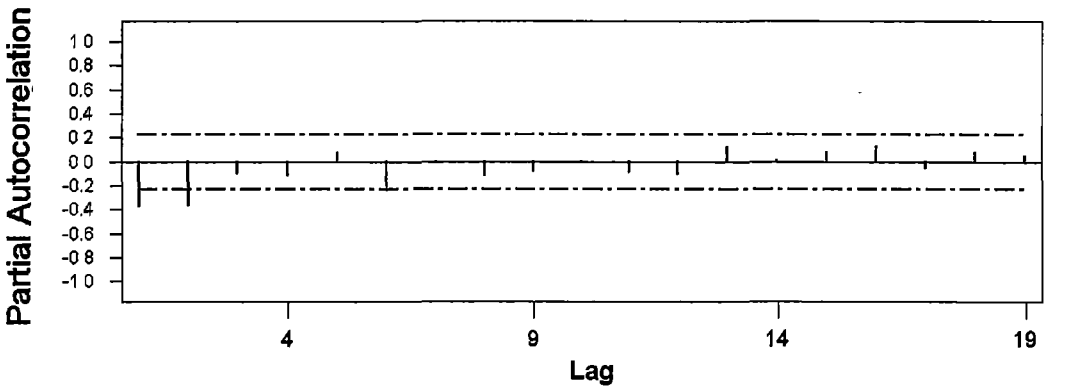


Fig 8.4 (8) : PACF Plot Of La139 After First Difference, Sample 43 (Plateau).

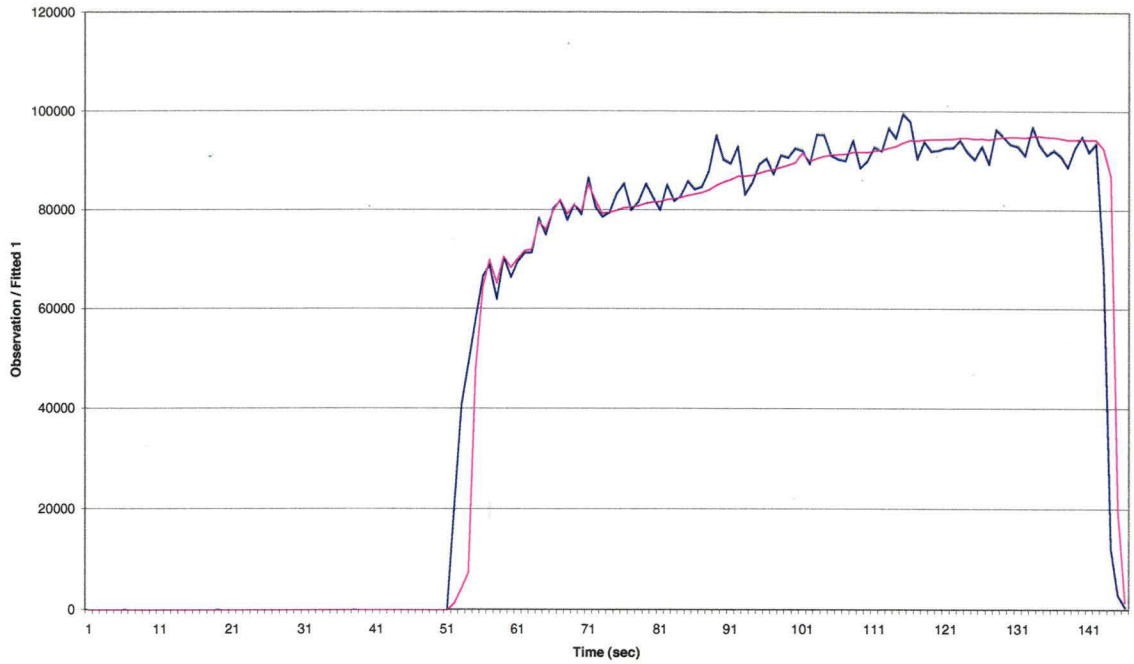


Fig 8.4 (9) : *Tab-Chart Of La139, Sample 43. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 43 in element La139, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

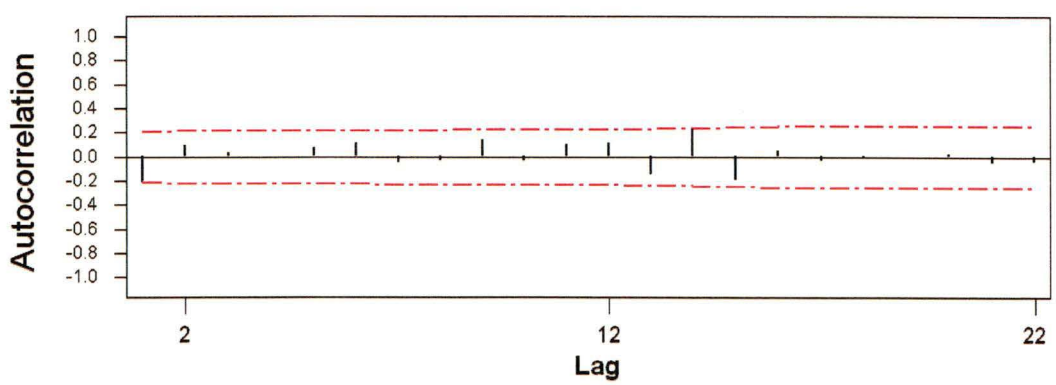


Fig 8.4 (10) : *ACF Plot Of Lu175 Sample 39 (Plateau). There is no significance spike in ACF*

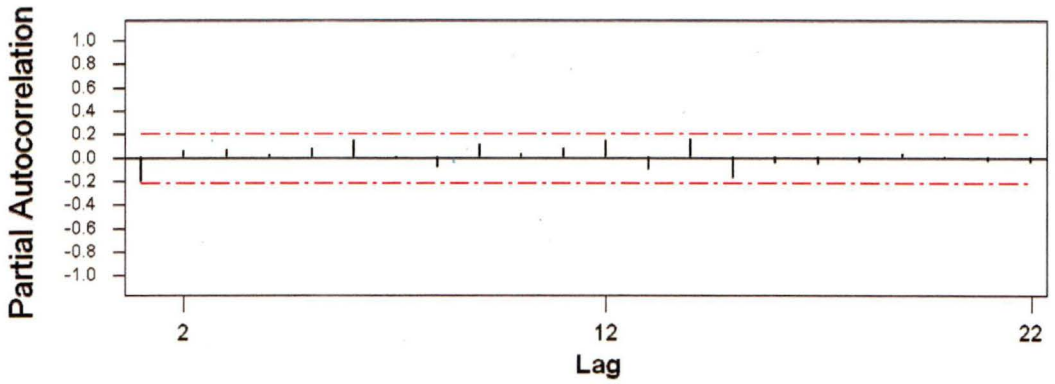


Fig 8.4 (11) : *PACF Plot Of Lu175, Sample 39 (Plateau). There is no significance spike in PACF*

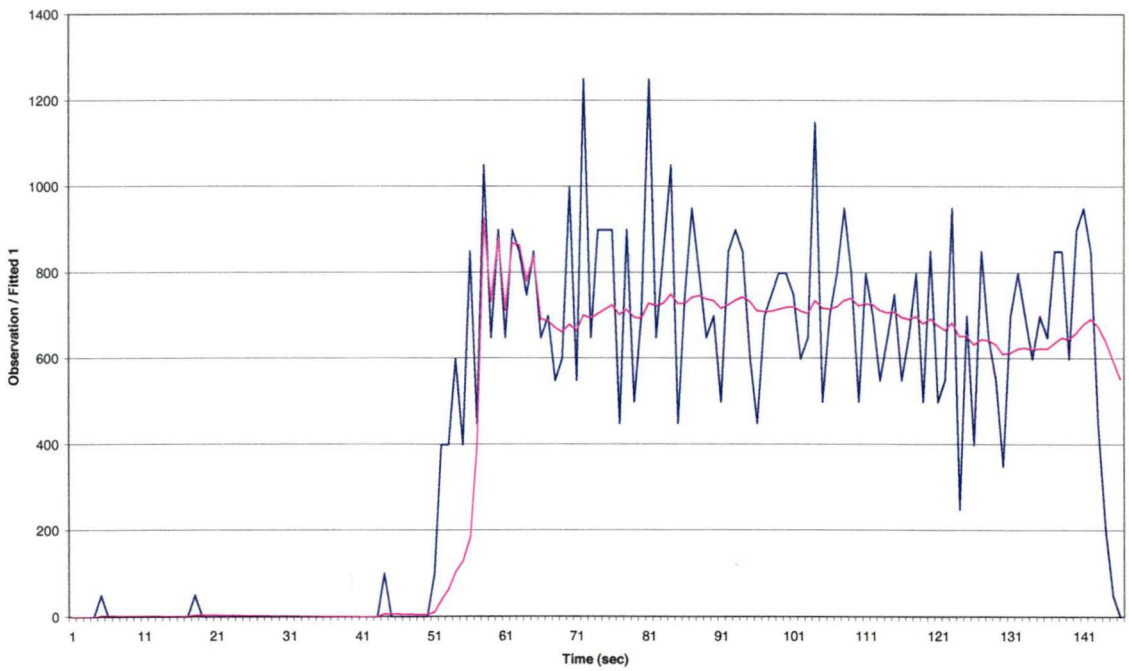


Fig 8.4 (12) : *Tab-Chart Of Lu175, Sample 39. The spiky curve is the tab-chart (i.e. actual observation series) of secondary standard sample 39 in element Lu175, whereas the smooth curve is the HOLT model (i.e. the fitted value (a) of modified linear holt exponential smoothing)*

(8.5) Comparison Of HOLT Model And ARIMA Model

ARIMA model is complicated and the users should have considerable statistical knowledge. Also it will involve subjective judgement, so that it seems to be not a practical method for automated use. It is not suitable and workable in this project because the researchers are looking for an automatic, fast and objective method to extract the information from the tab-chart. Thus ARIMA model does not match the aims and goals of this project (see section 1.10).

Exponential smoothing method or HOLT model is more user-friendly and automatic than ARIMA model. It requires only the formula calculations (i.e. calculation of the level, trend and seasonal components in the data) and parameters tunning (see Chapter 5). This method is relatively simple and objective and automatic.

(8.6) Chapter Conclusions

After fitting the ARIMA model to all the tab-charts, we obtain the following the analysis results. Over the background, most of the tab-charts are ARMA(1,1), random process and inconclusive. Over the plateau, most of the stationary tab-charts are AR(1) or ARMA(1,1) and most of the non-stationary tab-charts are ARIMA(0,1,1). The above results imply that a lot of tab-charts are autocorrelation. For the inconclusive tab-charts, statistical tests (eg. Durbin-Watson statistic) are needed to check that they are random process or autocorrelation in the further studies.

Since the property of ARIMA does not meet the requirement of the researchers, the ARIMA is not the main theme of this thesis. However, the ARIMA model can play the role of pre-processing the tab-chart (i.e. check the autocorrelation of tab-chart).

Chapter 9 Further Studies

(9.1) Nonparametric Regression

Apart from the parametric model (eg linear holt exponential smoothing model), the other non-parametric fitting (smoothing) models (eg. locally weighted regression) will also be explored and investigated for fitting the tab-chart in the next step. The purpose is to find other better models to fit the tab-chart such that more information can be extracted from the tab-chart. As mentioned in Chapter 5, the linear Holt exponential smoothing model allows the fast-past to still play a part in estimates. However, locally weighted regression considers the neighbouring points and ignoring the faraway points by introducing a weight function.

(9.2) Unknown Sample

After the primary and secondary standards are studied thoroughly by fitting models, this project will move to study the unknown samples (i.e. real samples) and try to apply the existing and explored methods to the unknown samples. The performance will be measured and make any necessary modifications, so that the models will be more suitable for the unknown sample.

(9.3) R-Software

The researchers currently use the Microsoft-Excel for data processing and calculation. This project will migrate from Microsoft-Excel to other statistical packages (eg. R-software). These softwares are designed and specialized in statistical analysis so that it is helpful to continue the tab-chart studies. If possible, these statistical packages are built into the machine and replace the Microsoft-Excel.

(9.4) Classification Rules

The classifications in the secondary samples are relatively approximated and are made by observation of the plots and results. Thus more advanced mathematical techniques should be employed to devise better classification rules, eg. discrimination analysis and neural network.

(9.5) Classification Of Spikes

As mentioned in the previous chapter, the researchers identify the spikes of the tab-chart by simply observations. Thus one of this project tasks is to find possible methods to classify these spikes in systematic and automatic way. Therefore, some statistical tests may be studies and used to classify the spikes of the tab-charts into real signal (i.e. caused by the mineral existence) or noise signal.

(9.6) Minerals Studied

In this thesis, the elements in samples are studied individually and separately. In other words, tab-chart of each element in a sample is studied individually and separately. However, the mineral in the real sample is a chemical compound of elements. Therefore, the change of composition in real samples will involve the investigation of more than one tab-chart.

References

1. O. D. Anderson (1976), Time Series Analysis and Forecasting : The Box-Jenkins Approach, London : Butterworth and Co (Publishers) Ltd, vii + 182pp.

2. George E. P. Box (1976), Gwilym M. Jenkins, Time Series Analysis : Forecasting And Control, San Francisco : Holden-day, xxi + 575pp.

3. Bruce L. Bowerman, Richard T. O'Connell and David A. Dickey (1986), Linear Statistical Models : An Applied Approach, Boston : Duxbury Press, ix + 690pp.

4. David C. Carslaw, Karl Ropkins and Margaret C. Bell (2006), "Change-point detection of gaseous and particulate traffic-related pollutants at a roadside location", Environmental Science and Technology, Vol 40, No 22, 6912-6918.

5. Chien-Hsiang Chen, "Segmentation of time series with clustering and genetic algorithm", National Cheng Kung University, Taiwan, http://etdncku.lib.ncku.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0727106-172352. Date accessed : 22, Jan 2008.

6. William S. Cleveland (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots", Journal of the American Statistical Association, Vol 74, No 368, 829-836.

7. Tak-chung Fu, Fu-lai Chung and Chak-man Ng, "Financial time series segmentation based on specialized binary tree representation", Proceedings of the 2006 International Conference on Data Mining, 3-9.

8. Wayne A. Fuller (1996), Introduction to Statistical Time Series, New York : John Wiley and Sons, xxii + 698pp.

9. Sarah Gilbert, Leonid Danyushevsky, "Laser Ablation ICPMS Methodology", Unpublished Technical Note, Centre for Ore Deposit Research, University of Tasmania.

10. John C. Hoff (1983), A Practical Guide To Box-Jenkins Forecasting, Belmont, Calif. : Lifetime Learning Publications, xii + 316pp.

11. P. Hubert (2000), "The segmentation procedure as a tool for discrete modeling of hydrometeorological regimes", Stochastic Environmental Research and Risk Assessment, Vol 14, No 4, 297-304.

12. B. H. Juang, Stephen E. Levinson (1986), "Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chain", IEEE Transaction on Information Theory, Vol IT-32, No 2, pp307-309.

13. Kuldeep Kumar and Berlin Wu (2001), "Detection of change point in time series analysis with fuzzy statistics", International Journal of System Science, Vol 32, No 9, 1185-1192.

14. Fang Li, George C. Runger and Eugene TUV (2006), "Supervised learning for change-point detection", International Journal of Production Research, Vol 44, No 14, 2853-2868.

15. Louis A. Liporace (1982), "Maximum Likelihood Estimation for Multivariate Observations of Markov Sources", IEEE Transactions on Information Theory, Vol IT-28, No 5, pp729-734.

16. Clive R. Neal, "Geochemical analysis of small samples: Micro-analytical techniques for the nineties and beyond", http://www.agu.org/revgeophys/neal00/neal00.html. Date accessed : 04, Sept, 2007.

17. John Neter, William Wasserman and Michael H. Kutner (1989), Applied Linear Regression Models, Homewood, III : Richard D. IRWIN, Inc, xv + 667pp.

18. M. B. Priestley (1981), Spectral Analysis and Time series, Volume 1 : Univariate Series, London : Academic Press, xviii + 653pp.

19. M. B. Priestley (1981), Spectral Analysis and Time series, Volume 2 : Multivariate series, Prediction and Control, London : Academic Press, xviii + 237pp.

20. Lawrence R. Rabiner (1989), "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, Vol 77, No. 2, pp257-286.

21. E. Schechtman, G. Bandner and S. Meginy (2007), "Detecting a change in a scale parameter – a combination of SPC and change point procedures", International Journal of Production Research, Vol 45, No 23, 5535-5545.

22. Paul J. Sylvester (2005), "Laser Ablation ICP-MS Developments and Trends for 2003", Geostandards and Geoanalytical Research, Vol 29, Issue 1, pp41-52.

23. Paul J. Sylvester (2006), "Trends in Analytical Developments and Earth Science Applications in LA-ICP-MS and LA-MC-ICP-MS for 2004 and 2005", Geostandards and Geoanalytical Research, Vol 30, Issue 3, pp197-207.

24. Wayne A. Taylor (2000), "Change-point analysis : a powerful new tool for detecting changes", http://www.variation.com/cpa/tech/changepoint.html. Date accessed : 22, Jan 2008.

25. R. John Watling, Allen Thomas, "Western Australian Inductively Coupled Plasma – Mass Spectrometric Facility", http://www.jdlcms.curtin.edu.au/icp_ms.html. Date accessed : 04, Sept 2007.

26. Wikipedia, http://en.wikipedia.org/wiki/Laser_ablation. Date accessed : 21 August 2007.

27. Wikipedia, http://en.wikipedia.org/wiki/Exponential_smoothing. Data accessed : 31 October 2007.

Appendix A : Notations And Formulas

(A1) Notations And Definitions

(A1.1) Notations For Concentration Calculation

$(Y)_B^A$ = the value Y of element A in sample B

where

$$A = \begin{cases} IE & \text{for internal standard element} \\ E & \text{for element E} \end{cases}$$

and

$$B = \begin{cases} US(t) & \text{for unknown sample t} \\ STD(t) & \text{for standard sample t} \\ STD & \text{for all standard samples} \end{cases}$$

Note :

$(Y)_{STD(0)}^A$ = known value Y of element A in standard sample

(A1.2) Notations For Detection Limit (DL) Calculation

(a) σ_{BG} = standard deviation of signal intensity over integration interval of background

(b) σ_{SG} = standard deviation of the signal intensity over integration interval of plateau

(c) n_{BG} = total number of data over the integration interval of background

(d) n_{SG} = total number of data over the integration interval of plateau

(A1.3) Definition For Total Analytical Error

$$BG_{Error} = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}} \quad \text{and} \quad SG_{Error} = \frac{\sigma_{SG}}{\sqrt{n_{SG}}}$$

Note : The notations BG_{Error} and SG_{Error} come from the unpublished technical paper [9]. In this thesis, these two notations are replaced by $SE_2(BG)$ and SE(PL) respectively.

(A2) Formulas

(A2.1) Formula Of Adjusted Average Counts Per Second Of Plateau

$$\overline{cps}_{\Delta SGBG} = \overline{cps}_{SG} - \overline{cps}_{BG}$$

where

$\overline{cps}_{BG}$ = average count per second (cps) over integration intervals for background

$\overline{cps}_{SG}$ = average count per second (cps) over integration intervals for plateau

(A2.2) Machine Drift

$$X_{Pre} = \frac{1}{m}\sum_{t=1}^{m}\left(\frac{\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)^{E}_{STD(t)}}{\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)^{IE}_{STD(t)}}\right)$$

$$X_{Post} = \frac{1}{m}\sum_{t=m+1}^{2m}\left(\frac{\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)^{E}_{STD(t)}}{\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)^{IE}_{STD(t)}}\right)$$

$$(D)^{E}_{US(t)} = X_{Post} + (X_{Pre} - X_{Post}) \times \frac{\text{Total Unknown Sample Analysis} + m - \text{Unknown Sample } t - 0.5}{\text{Total Unknown Sample Analysis} + m}$$

where

(a) m = total number of standard samples are analysed before or after unknown samples

(b) D = the drift of the machine

(c) $\left(\frac{\overline{cps}_{\Delta SGBG}}{ppm}\right)^{A}_{STD(t)}$ can be calculated by $\frac{(\overline{cps}_{\Delta SGBG})^{A}_{STD(t)}}{(ppm)^{A}_{STD(0)}}$

Note :

$$(D)^{IE}_{STD(t)} = (D)^{IE}_{US(t)} = 1 \qquad \text{since} \qquad X_{Pre} = X_{Post} = 1$$

(A2.3) ***Formula Of Calculating Concentration***

$$(\text{ppm})^{E}_{US(t)} = \frac{(\overline{\text{cps}}_{\Delta SGBG})^{E}_{US(t)}}{(D)^{E}_{US(t)} \times \left(\frac{\overline{\text{cps}}_{\Delta SGBG}}{\text{ppm}} \right)^{IE}_{US(t)}}$$

where

(a) ppm = concentration

(b) $\left(\frac{\overline{\text{cps}}_{\Delta SGBG}}{\text{ppm}} \right)^{IE}_{US(t)}$ can be calculated by $\frac{(\overline{\text{cps}}_{\Delta SGBG})^{IE}_{US(t)}}{(\text{ppm})^{IE}_{STD(0)}}$

Note :

$$(\text{ppm})^{IE}_{US(t)} = (\text{ppm})^{IE}_{STD(t)} = (\text{ppm})^{IE}_{STD(0)}$$

(A2.4) ***Formula Of Calculating Detection Limit (DL)***

$$(DL)^{E}_{US(t)} = \frac{3 \times (\sigma_{BG})^{E}_{US(t)} \times \sqrt{\frac{1}{(n_{BG})^{E}_{US(t)}} + \frac{1}{(n_{SG})^{E}_{US(t)}}}}{(D)^{E}_{US(t)} \times \left(\frac{\overline{\text{cps}}_{\Delta SGBG}}{\text{ppm}} \right)^{IE}_{US(t)}}$$

where

(a) DL = detection limit

(b) $\left(\frac{\overline{\text{cps}}_{\Delta SGBG}}{\text{ppm}} \right)^{IE}_{US(t)}$ can be calculated by $\frac{(\overline{\text{cps}}_{\Delta SGBG})^{IE}_{US(t)}}{(\text{ppm})^{IE}_{STD(0)}}$

(A2.5) ***Formula Of Calculating Total Analytical Error***

$$(AE)^{E}_{US(t)} = 100 \times \sqrt{\{(E)^{E}_{US(t)}\}^2 + \{(E)^{IE}_{US(t)}\}^2 + \{(E)^{E}_{STD}\}^2 + \{(E)^{IE}_{STD}\}^2}$$

where

(a) AE = total analytical error

(b) $$(E)^{E}_{US(t)} = \frac{\sqrt{(BG_{Error})^{E}_{US(t)} \times (BG_{Error})^{E}_{US(t)} + (SG_{Error})^{E}_{US(t)} \times (SG_{Error})^{E}_{US(t)}}}{(\overline{cps}_{\Delta SGBG})^{E}_{US(t)}}$$

(c) $$(E)^{IE}_{US(t)} = \frac{\sqrt{(BG_{Error})^{IE}_{US(t)} \times (BG_{Error})^{IE}_{US(t)} + (SG_{Error})^{IE}_{US(t)} \times (SG_{Error})^{IE}_{US(t)}}}{(\overline{cps}_{\Delta SGBG})^{IE}_{US(t)}}$$

(d) $$(E)^{E}_{STD} = \frac{1}{(\overline{cps}_{\Delta SGBG})^{E}_{STD}} \times \frac{1}{2m} \sum_{t=1}^{2m} \sqrt{(BG_{Error})^{E}_{STD(t)} \times (BG_{Error})^{E}_{STD(t)} + (SG_{Error})^{E}_{STD(t)} \times (SG_{Error})^{E}_{STD(t)}}$$

(e) $$(E)^{IE}_{STD} = \frac{1}{(\overline{cps}_{\Delta SGBG})^{IE}_{STD}} \times \frac{1}{2m} \sum_{t=1}^{2m} \sqrt{(BG_{Error})^{IE}_{STD(t)} \times (BG_{Error})^{IE}_{STD(t)} + (SG_{Error})^{IE}_{STD(t)} \times (SG_{Error})^{IE}_{STD(t)}}$$

Note :

(a) $$(\overline{cps}_{\Delta SGBG})^{E}_{STD} = (D)^{E}_{US(t)} \times \frac{1}{2m} \left(\sum_{t=1}^{2m} \left(\frac{\overline{cps}_{\Delta SGBG}}{ppm} \right)^{IE}_{STD(t)} \right) \times (ppm)^{E}_{STD(0)}$$

(b) $$(\overline{cps}_{\Delta SGBG})^{IE}_{STD} = \frac{1}{2m} \left(\sum_{t=1}^{2m} \left(\frac{\overline{cps}_{\Delta SGBG}}{ppm} \right)^{IE}_{STD(t)} \right) \times (ppm)^{IE}_{STD(0)}$$

(c) The total analytical error of a standard sample (i.e. $(AE)^{A}_{STD(t)}$) is analytical noise

(d) The total analytical error of a unknown sample (i.e. $(AE)^{A}_{US(t)}$) is the sample variability

Appendix B : ARIMA Result

(in the attached CD)

Appendix C : Raw Data

(in the attached CD)

Appendix D : Result Of Classification Rules

(in the attached CD)

Appendix E : Performance Table Of The Classification Rules

(E1) Notations And Explanation Of The Tables

(1) Column of the tables represents the element and row of the tables represents the sample

(2) Number "0" in the tables means the classification rules is not good in this sample

(3) Number "1" in the tables means the classification rules is good in this sample

(4) The number in red colour means that the background and plateau of the tab-chart are very close

(5) The last row of tables (i.e. "*Succ* =") means the successful rate of classification rules in the element

(6) Ca43 is special and different from other elements since the tab-chart of this element has two plateaus

(E2) Performance Table Of The Classification Rules (Table 1)

Sample	Ba137	Ca43	Ce140	Cu65	Dy163	Er166	Eu153	Gd157	Hf178	La139	Lu175	Nb93	Nd146	Ni60
1	0	1	0	1	0	1	1	0	1	1	1	1	1	1
2	0	0	1	0	1	1	1	1	1	1	1	0	0	1
3	1	0	1	1	0	1	1	1	1	1	0	0	0	1
4	1	0	1	0	1	1	1	1	0	1	1	1	1	0
5	1	0	1	1	1	1	0	1	1	1	1	0	1	1
6	0	0	1	1	1	0	1	1	1	0	1	1	0	1
7	1	1	1	1	0	1	0	1	1	1	1	1	1	1
8	1	1	0	0	0	0	1	1	1	1	1	1	1	1
9	1	0	0	0	1	0	0	0	1	0	0	1	0	1
10	0	1	0	1	0	1	1	0	1	1	0	1	1	1
11	0	1	1	1	0	0	1	1	1	1	0	0	0	0
12	1	1	0	0	0	1	1	0	0	0	0	1	1	1
13	1	0	1	1	0	1	0	1	0	1	0	0	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	0	0	0	0	0	0	0	0	1	0	0
16	1	1	0	0	0	0	0	0	0	1	0	1	1	0
17	0	0	0	0	0	0	0	0	0	1	0	1	0	0
18	0	1	1	0	0	0	0	0	0	0	0	0	1	0
19	1	1	1	1	0	0	0	0	0	1	0	0	1	0
20	1	0	0	0	0	0	0	0	1	1	0	0	1	0
21	1	1	1	0	0	0	0	0	0	1	0	0	1	0
22	0	1	0	0	1	0	0	0	0	1	0	1	1	0
23	1	1	1	1	0	0	0	0	0	1	0	1	1	0
24	1	0	1	0	0	1	0	1	0	0	0	0	1	0
25	1	0	1	1	0	0	0	0	1	1	0	1	1	0
26	1	0	1	1	0	0	0	1	1	1	0	1	1	0
27	1	1	1	0	0	1	1	1	1	1	1	1	1	1
28	0	1	0	0	1	1	1	1	1	1	0	0	1	1
29	0	1	1	0	0	1	1	1	1	1	1	1	1	1
30	1	1	1	1	1	1	1	1	1	1	1	1	0	1
31	1	1	1	1	0	1	1	0	1	1	1	1	1	0
32	1	1	1	1	0	1	1	1	1	0	1	1	0	1
33	1	1	1	0	1	1	1	1	1	1	1	0	1	1
34	1	0	0	0	1	1	0	1	0	1	1	1	1	1
35	1	1	0	1	0	0	1	1	1	1	1	1	1	1
36	1	1	1	0	1	1	1	1	1	0	1	1	1	0
37	1	1	1	1	1	0	1	1	1	1	1	1	1	1
38	0	0	1	0	1	1	1	1	1	1	1	1	1	1
39	1	1	1	0	1	1	1	1	1	0	1	1	1	0
40	1	1	0	1	1	1	1	1	1	1	1	1	1	1
41	1	0	1	1	1	1	1	1	1	1	1	1	1	1
42	1	1	1	1	0	1	1	0	1	1	1	1	1	1
43	1	1	0	1	1	1	0	1	1	1	0	1	1	1
44	0	1	1	1	1	1	1	1	1	1	1	1	1	1
45	1	1	1	0	0	0	0	0	0	1	0	1	1	0
46	1	1	1	0	0	0	0	0	0	1	0	1	1	0
47	1	0	1	0	0	1	1	0	1	1	0	1	1	1
48	1	1	1	1	0	0	1	1	0	1	0	1	1	0
49	0	1	1	0	0	0	0	0	0	1	0	0	1	0
50	1	1	0	0	0	0	0	0	0	1	0	1	0	0
51	0	1	0	0	0	0	0	0	0	0	0	1	0	0
52	0	0	1	0	0	0	0	0	0	1	0	1	1	0
53	1	1	0	1	0	0	1	1	1	1	0	1	0	1
54	1	1	1	0	0	0	1	1	1	1	0	1	1	0
55	1	0	1	0	0	1	1	1	1	1	1	1	1	1
56	0	1	1	1	1	1	1	1	1	0	0	1	1	1
57	1	1	0	1	1	1	1	1	1	1	1	1	1	1
58	1	0	1	0	1	1	1	1	1	1	1	1	1	1
59	1	1	1	1	1	1	1	1	1	1	0	1	1	1
60	0	1	0	1	1	0	1	0	1	1	1	1	1	1
61	1	1	1	0	1	1	1	1	0	1	1	1	1	1
62	0	0	1	1	1	1	0	1	1	1	1	1	1	0
63	0	1	1	0	1	1	1	1	1	1	0	1	1	0
64	1	1	1	0	1	1	1	1	0	1	1	1	1	1
65	1	1	0	0	1	1	1	0	1	1	1	1	1	1
66	0	0	1	0	1	1	1	1	1	1	1	1	1	1
Succ =	0.69697	0.681818	0.69697	0.454545	0.454545	0.590909	0.621212	0.621212	0.666667	0.833333	0.5	0.80303	0.818182	0.590909

(E3) Performance Table Of The Classification Rules (Table 2)

Sample	*Pb208*	*Rb85*	*Sc45*	*Sm147*	*Sr88*	*Ta181*	*Th232*	*Ti47*	*U238*	*V51*	*Y89*	*Yb172*	*Zn66*	*Zr90*
1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
2	0	1	1	1	1	0	1	0	1	1	1	1	0	1
3	1	1	0	1	0	1	0	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	0	1	0	0	0
5	1	1	1	1	1	1	1	1	1	1	1	1	0	1
6	1	1	1	1	0	0	0	1	1	0	1	1	1	0
7	1	1	1	0	1	1	0	1	1	1	1	1	1	1
8	1	0	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	0	1	0	1	1	1	1	1	0	0	0
10	1	0	0	1	1	1	1	1	1	0	1	1	0	1
11	1	1	0	0	1	0	1	0	1	1	1	1	1	1
12	1	1	0	1	1	1	1	1	1	1	1	0	1	1
13	0	1	0	1	1	0	1	1	1	0	1	1	1	1
14	0	1	1	1	1	1	1	0	1	0	0	1	0	1
15	1	1	1	*0*	1	*0*	0	0	*0*	0	1	*0*	1	1
16	1	1	0	*0*	1	*0*	1	0	0	1	1	*0*	1	1
17	1	1	1	*0*	0	*0*	1	1	*0*	1	1	*0*	1	1
18	1	1	1	0	1	*0*	0	1	0	1	1	0	0	1
19	1	0	0	*0*	0	*0*	0	0	0	1	0	*0*	1	1
20	1	1	0	0	1	*0*	1	0	1	1	0	0	1	0
21	1	1	1	*0*	0	*0*	0	1	1	1	1	*0*	1	1
22	1	1	0	0	1	*0*	0	0	1	1	1	0	1	1
23	0	0	1	*0*	0	*0*	1	1	0	1	0	0	0	1
24	1	1	0	0	0	0	1	1	0	1	1	0	1	1
25	0	1	1	0	0	0	0	1	0	0	1	1	1	1
26	1	1	0	0	1	0	1	1	1	0	1	0	1	1
27	1	1	0	1	1	1	1	0	1	1	1	1	1	1
28	1	1	0	1	0	1	1	1	1	1	1	1	1	1
29	0	0	1	1	1	1	1	1	1	1	1	1	1	1
30	1	0	1	1	1	1	1	1	1	0	1	1	0	1
31	1	0	1	1	0	1	0	1	1	0	0	1	1	0
32	1	1	1	1	1	1	0	1	0	0	1	1	1	1
33	1	1	1	1	1	1	1	1	1	0	1	1	1	1
34	1	0	1	1	1	1	0	0	1	1	1	1	1	1
35	1	0	0	1	1	1	1	0	0	1	1	1	1	1
36	0	0	1	1	1	1	1	1	0	1	1	1	1	1
37	1	1	1	1	0	1	1	1	1	0	1	0	1	1
38	1	1	0	0	1	0	1	0	1	1	1	1	1	1
39	1	0	1	1	1	1	1	1	0	1	1	1	1	1
40	1	0	1	1	1	0	1	0	0	0	0	1	1	1
41	0	1	1	1	1	1	1	1	0	1	1	1	0	1
42	1	0	1	1	1	1	1	1	1	1	1	1	0	1
43	1	1	1	1	1	0	1	1	1	0	1	1	1	1
44	1	1	1	0	1	1	1	0	1	1	0	1	0	1
45	0	1	0	*0*	1	*0*	1	1	0	1	0	0	0	1
46	1	1	0	*0*	1	*0*	1	0	0	1	1	*0*	0	1
47	1	0	1	1	1	*0*	1	0	1	0	1	0	0	0
48	0	0	0	1	1	0	1	0	1	0	1	0	1	1
49	0	0	0	*0*	0	*0*	0	1	0	1	1	*0*	0	1
50	1	1	1	*0*	0	*0*	0	1	*0*	1	0	*0*	0	0
51	0	1	1	0	1	0	1	0	1	1	1	*0*	1	1
52	1	1	1	*0*	1	*0*	0	0	0	1	1	*0*	1	1
53	1	0	1	0	1	0	1	1	1	1	1	0	1	1
54	1	1	0	0	1	0	0	1	0	0	0	0	1	1
55	0	1	1	0	1	0	1	0	1	0	1	1	1	1
56	1	0	0	1	0	1	1	0	1	1	0	1	1	1
57	1	1	0	1	1	1	1	0	1	0	1	1	1	1
58	1	0	0	1	0	1	1	1	1	1	1	1	1	1
59	1	0	0	1	0	1	1	1	1	1	1	1	1	1
60	0	0	1	1	1	1	0	1	0	1	1	1	0	1
61	1	1	0	0	0	1	1	1	1	1	1	1	1	1
62	0	1	0	1	1	1	1	0	1	1	1	1	1	1
63	1	0	1	0	1	1	1	1	1	0	1	1	1	1
64	1	1	0	0	1	1	0	1	1	1	1	1	0	1
65	0	1	1	1	1	1	1	0	1	1	1	1	1	1
66	1	1	1	1	1	0	1	0	1	1	1	1	1	1
Succ =	*0.757576*	*0.666667*	*0.590909*	*0.560606*	*0.742424*	*0.515152*	*0.712121*	*0.606061*	*0.681818*	*0.681818*	*0.833333*	*0.621212*	*0.712121*	*0.893939*

Appendix F : Statistical Summaries In Classification Method

(F1) Notations

(1) n_{BG} = the length of background determined by intuitive method or classification method

(2) n_{SG} = the length of plateau determined by intuitive method or classification method

(3) $\bar{x}_{BG}$ = mean of background determined by intuitive method or classification method

(4) $\bar{x}_{SG}$ = mean of plateau determined by intuitive method or classification method

(5) σ_{BG} = SD of background determined by intuitive method or classification method

(6) σ_{SG} = SD of plateau determined by intuitive method or classification method

(7) $SE_1(BG)$ = standard error of background mean

(8) SE(PL) = standard error of plateau mean

(9) SD = standard deviation

(10) The fitted value L_t and sample 5 in Th232 are used in this example

(F2) Table Of Classification Method's Results (Sample 5, Element Th232)

Obs	*Count per second*	*Fitted Value (L_d)*	*Result*	*Obs*	*Count per second*	*Fitted Value (L_d)*	*Result*
1	0.00	0.00	Background	*61*	7651.31	7707.45	Jump
2	0.00	0.00	Background	*62*	7401.22	7544.64	Jump
3	0.00	0.00	Background	*63*	6600.97	6848.93	Jump
4	0.00	0.00	Background	*64*	8101.46	7870.50	Jump
5	0.00	0.00	Background	*65*	8301.54	8271.84	Jump
6	0.00	0.00	Background	*66*	8401.57	8436.88	Jump
7	0.00	0.00	Background	*67*	7551.27	7784.00	Jump
8	0.00	0.00	Background	*68*	7051.11	7216.06	Jump
9	0.00	0.00	Background	*69*	6901.06	7156.71	Plateau
10	0.00	0.00	Background	*70*	6450.93	7069.29	Plateau
11	50.00	3.50	Background	*71*	7901.39	7087.79	Plateau
12	0.00	3.39	Background	*72*	8201.50	7128.28	Plateau
13	50.00	6.77	Background	*73*	7451.24	7116.42	Plateau
14	0.00	6.54	Background	*74*	7701.32	7123.84	Plateau
15	0.00	6.30	Background	*75*	8101.46	7160.36	Plateau
16	0.00	6.07	Background	*76*	8401.57	7217.97	Plateau
17	0.00	5.83	Background	*77*	7951.41	7243.35	Plateau
18	0.00	5.60	Background	*78*	7851.38	7261.94	Plateau
19	0.00	5.36	Background	*79*	7351.21	7245.86	Plateau
20	0.00	5.13	Background	*80*	6400.91	7164.68	Plateau
21	0.00	4.90	Background	*81*	7501.25	7164.07	Plateau
22	0.00	4.67	Background	*82*	6600.97	7101.42	Plateau
23	0.00	4.44	Background	*83*	9151.87	7220.33	Plateau
24	0.00	4.22	Background	*84*	9451.99	7357.32	Plateau
25	0.00	4.00	Background	*85*	7401.22	7347.04	Plateau
26	0.00	3.78	Background	*86*	8251.52	7397.15	Plateau
27	100.00	10.57	Background	*87*	8251.52	7446.15	Plateau
28	0.00	10.14	Background	*88*	7801.36	7462.46	Plateau
29	0.00	9.71	Background	*89*	6300.88	7373.54	Plateau
30	0.00	9.28	Background	*90*	6500.94	7301.85	Plateau
31	0.00	8.85	Background	*91*	7251.17	7285.45	Plateau
32	0.00	8.43	Background	*92*	8001.43	7322.62	Plateau
33	100.00	15.01	Background	*93*	8451.59	7390.60	Plateau
34	0.00	14.38	Background	*94*	10102.28	7572.34	Plateau
35	0.00	13.74	Background	*95*	9652.08	7716.93	Plateau
36	0.00	13.12	Background	*96*	7201.16	7685.25	Plateau
37	0.00	12.49	Background	*97*	8051.45	7713.95	Plateau
38	0.00	11.88	Background	*98*	7901.39	7731.09	Plateau
39	0.00	11.28	Background	*99*	8051.45	7758.00	Plateau
40	0.00	10.69	Background	*100*	6650.99	7685.82	Plateau
41	0.00	10.11	Background	*101*	6150.84	7580.79	Plateau
42	50.00	13.04	Background	*102*	5900.78	7461.60	Plateau
43	0.00	12.37	Background	*103*	6951.08	7419.91	Plateau
44	0.00	11.71	Background	*104*	7351.21	7407.83	Plateau
45	0.00	11.07	Background	*105*	8001.43	7441.95	Plateau
46	50.00	13.94	Background	*106*	8401.57	7503.26	Plateau
47	0.00	13.21	Background	*107*	8851.75	7594.30	Plateau
48	50.00	15.99	Background	*108*	6951.08	7549.45	Plateau
49	0.00	15.18	Background	*109*	7501.25	7544.57	Plateau
50	2400.13	182.38	Background	*110*	8751.71	7627.45	Plateau
51	4100.37	463.11	Background	*111*	7801.36	7641.15	Plateau
52	4750.50	779.89	Background	*112*	7201.16	7612.32	Plateau
53	5700.72	4722.53	Jump	*113*	7701.32	7619.37	Plateau
54	5800.75	5747.59	Jump	*114*	7701.32	7626.16	Plateau
55	5650.71	5841.08	Jump	*115*	8051.45	7657.19	Plateau
56	6650.99	6629.54	Jump	*116*	8051.45	7687.15	Plateau
57	7101.12	7150.77	Jump	*117*	7551.27	7681.03	Plateau
58	6150.84	6486.85	Jump	*118*	8351.56	7730.99	Plateau
59	6550.96	6620.40	Jump	*119*	7801.36	7740.67	Plateau
60	7601.29	7476.26	Jump	*120*	7551.27	7732.34	Plateau

Obs	*Count per second*	*Fitted Value (L_f)*	*Result*
121	7801.36	7741.60	Plateau
122	7451.24	7725.86	Plateau
123	7601.29	7720.96	Plateau
124	6250.87	7621.54	Plateau
125	7701.32	7626.77	Plateau
126	8051.45	7656.36	Plateau
127	7151.14	7621.95	Plateau
128	8601.65	7690.18	Plateau
129	7251.17	7661.64	Plateau
130	7601.29	7658.47	Plateau
131	8501.61	7718.37	Plateau
132	7751.34	7723.76	Plateau
133	6650.99	7651.82	Plateau
134	7101.12	7613.63	Plateau
135	7251.17	7587.18	Plateau
136	8851.75	7673.68	Plateau
137	7401.22	7655.88	Plateau
138	6450.93	7572.10	Plateau
139	7501.25	7564.57	Plateau
140	7351.21	7546.88	Plateau
141	7901.39	7568.40	Plateau
142	3250.23	7263.76	Plateau
143	500.01	6776.69	Plateau
144	250.00	1548.63	Drop
145	0.00	95.24	Drop
146	0.00	-210.68	Drop

(F3) Statistical Calculations Of Classification Method And Intuitive Method

(F3.1) Intuitive Method

(a) Start point of background determined by researcher observation = 1

(b) Start point of plateau determined by researcher observation = 70

(c) End point of background determined by researcher observation = 44

(d) End point of plateau determined by researcher observation = 136

(e) $n_{BG} = (c) - (a) + 1 = 44 - 1 + 1 = 44$

(f) $n_{SG} = (d) - (b) + 1 = 136 - 70 + 1 = 67$

(g) $\bar{x}_{BG}$ = the mean of count per second from (a) to (c) = 8.89

(h) $\bar{x}_{SG}$ = the mean of count per second from (b) to (d) = 7696.19

(i) σ_{BG} = the standard deviation of count per second from (a) to (c) = 24.52

(j) σ_{SG} = the standard deviation of count per second from (b) to (d) = 827.58

(k) $$SE_1(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}}} = (i) \times \sqrt{\frac{1}{(e)}} = 24.52 \times \sqrt{\frac{1}{44}} = 3.70$$

(l) $$SE_2(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}} = (i) \times \sqrt{\frac{1}{(e)} + \frac{1}{(f)}} = 24.52 \times \sqrt{\frac{1}{44} + \frac{1}{67}} = 4.76$$

(m) $$SE(PL) = \sigma_{SG} \times \sqrt{\frac{1}{n_{SG}}} = (j) \times \sqrt{\frac{1}{(f)}} = 827.58 \times \sqrt{\frac{1}{67}} = 101.11$$

(F3.2) ***Classification Method Of HOLT Model***

(a) Start point of background determined by classification method = 1

(b) Start point of plateau determined by classification method = 69

(c) End point of background determined by classification method = 52

(d) End point of plateau determined by classification method = 143

(e) n_{BG} = (c) – (a) + 1 = 52 – 1 + 1 = 52

(f) n_{SG} = (d) – (b) + 1 = 143 – 69 + 1 = 75

(g) $\bar{x}_{BG}$ = the mean of fitted value from (a) to (c) = 34.27

(h) $\bar{x}_{SG}$ = the mean of fitted value from (b) to (d) = 7486.82

(i) σ_{BG} = the standard deviation of fitted value from (a) to (c) = 125.15

(j) σ_{SG} = the standard deviation of fitted value from (b) to (d) = 223.75

(k) $SE_1(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}}} = (i) \times \sqrt{\frac{1}{(e)}} = 125.15 \times \sqrt{\frac{1}{52}} = 17.35$

(l) $SE_2(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}} = (i) \times \sqrt{\frac{1}{(e)} + \frac{1}{(f)}} = 125.15 \times \sqrt{\frac{1}{52} + \frac{1}{75}} = 22.58$

(m) $SE(PL) = \sigma_{SG} \times \sqrt{\frac{1}{n_{SG}}} = (j) \times \sqrt{\frac{1}{(f)}} = 223.75 \times \sqrt{\frac{1}{75}} = 25.84$

Appendix G : Statistical Summaries In Classification Rules

(G1) Notations

(1) n_{BG} = the length of background determined by intuitive method or classification rules

(2) n_{SG} = the length of plateau determined by intuitive method or classification rules

(3) $\bar{x}_{BG}$ = mean of background determined by intuitive method or classification rules

(4) $\bar{x}_{SG}$ = mean of plateau determined by intuitive method or classification rules

(5) σ_{BG} = SD of background determined by intuitive method or classification rules

(6) σ_{SG} = SD of plateau determined by intuitive method or classification rules

(7) $SE_1(BG)$ = standard error of background mean

(8) SE(PL) = standard error of plateau mean

(9) SD = standard deviation

(10) The fitted value L_t and sample 10 in Ba137 are used in this example

(G2) Table Of Classification Rules' Results (Sample 10, Element Ba137)

Obs	*Count per second*	*Fitted Value (L_J)*	*Result*	*Obs*	*Count per second*	*Fitted Value (L_J)*	*Result*
1	0.00	0.00	Not Sure	*61*	26966.26	26862.51	Jump
2	0.00	0.00	Not Sure	*62*	26816.07	27072.86	Jump
3	0.00	0.00	Not Sure	*63*	25864.96	26312.95	Jump
4	50.00	25.00	Not Sure	*64*	26315.49	26449.72	Jump
5	0.00	25.58	Background	*65*	32623.80	31502.24	Jump
6	0.00	26.04	Background	*66*	28317.93	29247.50	Jump
7	0.00	26.40	Background	*67*	27917.43	28327.42	Jump
8	0.00	26.65	Background	*68*	25915.02	26475.88	Jump
9	0.00	26.82	Background	*69*	26615.84	26432.86	Plateau
10	0.00	26.90	Background	*70*	28468.12	26523.02	Plateau
11	0.00	26.90	Background	*71*	28818.57	26636.84	Plateau
12	0.00	26.82	Background	*72*	29970.09	26829.42	Plateau
13	0.00	26.68	Background	*73*	27567.00	26849.09	Plateau
14	0.00	26.46	Background	*74*	28317.93	26921.96	Plateau
15	0.00	26.19	Background	*75*	26365.54	26856.97	Plateau
16	0.00	25.87	Background	*76*	26315.49	26791.65	Plateau
17	0.00	25.50	Background	*77*	28267.87	26866.24	Plateau
18	0.00	25.08	Background	*78*	29068.90	26995.60	Plateau
19	0.00	24.62	Background	*79*	27617.06	27020.08	Plateau
20	0.00	24.12	Background	*80*	29219.09	27156.67	Plateau
21	0.00	23.59	Background	*81*	30871.31	27405.12	Plateau
22	50.00	26.53	Background	*82*	26215.36	27319.97	Plateau
23	0.00	25.83	Background	*83*	31772.58	27626.69	Plateau
24	50.00	28.61	Background	*84*	28317.93	27681.72	Plateau
25	0.00	27.75	Background	*85*	27416.81	27671.61	Plateau
26	0.00	26.88	Background	*86*	27767.24	27686.02	Plateau
27	0.00	25.99	Background	*87*	30821.24	27913.43	Plateau
28	0.00	25.09	Background	*88*	24513.44	27691.51	Plateau
29	0.00	24.18	Background	*89*	27767.24	27704.00	Plateau
30	0.00	23.27	Background	*90*	25464.50	27554.60	Plateau
31	0.00	22.36	Background	*91*	26015.13	27448.35	Plateau
32	0.00	21.45	Background	*92*	24963.94	27271.94	Plateau
33	0.00	20.54	Background	*93*	26015.13	27175.00	Plateau
34	0.00	19.64	Background	*94*	26015.13	27081.60	Plateau
35	0.00	18.75	Background	*95*	25114.10	26928.67	Plateau
36	0.00	17.86	Background	*96*	31622.36	27236.95	Plateau
37	0.00	16.99	Background	*97*	22961.79	26929.69	Plateau
38	0.00	16.13	Background	*98*	22361.18	26590.79	Plateau
39	0.00	15.29	Background	*99*	29469.42	26761.34	Plateau
40	0.00	14.47	Background	*100*	22060.88	26408.94	Plateau
41	0.00	13.66	Background	*101*	25614.67	26317.79	Plateau
42	0.00	12.86	Background	*102*	20509.41	25873.69	Plateau
43	0.00	12.09	Background	*103*	24763.72	25743.46	Plateau
44	0.00	11.34	Background	*104*	25013.99	25637.12	Plateau
45	0.00	10.61	Background	*105*	21610.44	25298.23	Plateau
46	0.00	9.90	Background	*106*	25814.90	25267.05	Plateau
47	0.00	9.21	Background	*107*	25013.99	25183.53	Plateau
48	0.00	8.55	Plateau	*108*	29118.96	25392.72	Plateau
49	0.00	7.91	Plateau	*109*	23262.10	25187.72	Plateau
50	0.00	7.29	Plateau	*110*	21560.39	24872.56	Plateau
51	0.00	1.44	Jump	*111*	23362.21	24696.32	Plateau
52	12603.55	10082.88	Drop	*112*	28668.38	24900.11	Plateau
53	17406.77	16345.05	Jump	*113*	24763.72	24826.85	Plateau
54	18557.70	18688.10	Jump	*114*	20609.50	24467.76	Plateau
55	21860.69	21778.24	Jump	*115*	21810.64	24207.08	Plateau
56	23612.47	23810.89	Jump	*116*	23262.10	24059.53	Plateau
57	25514.55	25707.33	Jump	*117*	25013.99	24042.72	Plateau
58	25064.05	25695.38	Jump	*118*	24413.32	23987.75	Plateau
59	26766.02	26953.55	Jump	*119*	25714.78	24028.93	Plateau
60	24413.32	25293.02	Jump	*120*	22611.43	23854.71	Plateau

Obs	*Count per second*	*Fitted Value (L_i)*	*Result*
121	22661.48	23692.71	Plateau
122	24363.27	23658.28	Plateau
123	23762.63	23586.19	Plateau
124	25064.05	23610.75	Plateau
125	22561.38	23462.47	Plateau
126	22361.18	23308.03	Plateau
127	23862.73	23266.85	Plateau
128	22461.28	23132.13	Plateau
129	22811.63	23029.48	Plateau
130	22711.53	22926.41	Plateau
131	24213.11	22935.05	Plateau
132	20459.36	22683.91	Plateau
133	23412.26	22650.82	Plateau
134	25814.90	22790.37	Plateau
135	24613.55	22844.52	Plateau
136	24263.17	22875.30	Plateau
137	25414.44	22988.41	Plateau
138	24963.94	23068.86	Plateau
139	21160.01	22882.70	Plateau
140	9502.01	21888.70	Plateau
141	1800.07	20390.45	Plateau
142	800.01	18875.03	Plateau
143	100.00	3813.13	Jump
144	250.00	326.65	Drop
145	150.00	-462.91	Drop
146	50.00	-602.75	Drop

(G3) Statistical Calculations Of Classification Rules And Intuitive Method

(G3.1) Intuitive Method

(a) Start point of background determined by researcher observation = 1

(b) Start point of plateau determined by researcher observation = 71

(c) End point of background determined by researcher observation = 47

(d) End point of plateau determined by researcher observation = 135

(e) $n_{BG} = (c) - (a) + 1 = 47 - 1 + 1 = 47$

(f) $n_{SG} = (d) - (b) + 1 = 135 - 71 + 1 = 65$

(g) $\bar{x}_{BG}$ = the mean of count per second from (a) to (c) = 3.125

(h) $\bar{x}_{SG}$ = the mean of count per second from (b) to (d) = 25408.55

(i) σ_{BG} = the standard deviation of count per second from (a) to (c) = 12.23

(j) σ_{SG} = the standard deviation of count per second from (b) to (d) = 2826.66

(k) $SE_1(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}}} = (i) \times \sqrt{\frac{1}{(e)}} = 12.23 \times \sqrt{\frac{1}{47}} = 1.78$

(l) $SE_2(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}} = (i) \times \sqrt{\frac{1}{(e)} + \frac{1}{(f)}} = 12.23 \times \sqrt{\frac{1}{47} + \frac{1}{65}} = 2.34$

(m) $SE(PL) = \sigma_{SG} \times \sqrt{\frac{1}{n_{SG}}} = (j) \times \sqrt{\frac{1}{(f)}} = 2826.66 \times \sqrt{\frac{1}{65}} = 350.60$

(G3.2) Classification Rules Of HOLT Model

(a) Observations are classified as background by classification rules : 5 to 47

(b) Observations are classified as plateau by classification rules : 48, 49, 50 and 69 to 142

(c) $n_{BG} = 47 - 5 + 1 = 43$

(d) $n_{SG} = 1 + 1 + 1 + (142 - 69 + 1) = 77$

(e) $\bar{x}_{BG}$ = the mean of fitted value from (a) to (c) = 21.80

(f) $\bar{x}_{SG}$ = the mean of fitted value from (b) to (d) = 24238.77

(g) σ_{BG} = the standard deviation of fitted value in observations (a) = 5.86

(h) σ_{SG} = the standard deviation of fitted value in observations (b) = 5289.32

(i) $SE_1(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}}} = (g) \times \sqrt{\frac{1}{(c)}} = 5.86 \times \sqrt{\frac{1}{43}} = 0.89$

(j) $SE_2(BG) = \sigma_{BG} \times \sqrt{\frac{1}{n_{BG}} + \frac{1}{n_{SG}}} = (g) \times \sqrt{\frac{1}{(c)} + \frac{1}{(d)}} = 5.86 \times \sqrt{\frac{1}{43} + \frac{1}{77}} = 1.12$

(k) $SE(PL) = \sigma_{SG} \times \sqrt{\frac{1}{n_{SG}}} = (h) \times \sqrt{\frac{1}{(d)}} = 5289.32 \times \sqrt{\frac{1}{77}} = 602.77$

Appendix H : Tables Of Mean After Applying Classification Method

(in the attached CD)

Appendix I : Tables Of Standard Deviation After Applying Classification Method

(in the attached CD)

Appendix J : Terminology

(J) Terminology Table

	Term	Meaning
1.	Ar and He gas	This gas is responsible for carrying the ablated material from Nu wave UP-213 to the plasma of the ICP-MS instrument.
2.	ARIMA model	This time series model is a famous method in data fitting and forecasting. This model is proposed by Box and Jenkins in 1967.
3.	Amount	The actual amount of the element in the sample.
4.	Autocorrelation	The data in the time series is not independent.
5.	Background signal (noise)	It is the elementary signal even though the laser is off and no ablated material in the gas.
6.	Change point analysis	This is a problem in the data mining of detecting the mean shift or variance change in the time series.
7.	Classification method	It is used to divide the tab-chart into different portions (detection of change points in the tab-chart) by the human adjustment of trend, smooth parameters and threshold value (T).
8.	Classification rules	Similar to the classification method, it is used to divide the tab-chart into different portions. The difference is that these classification rules do the division automatically, not by human judgement.
9.	CODES	It stands for ARC Centre of Excellence in Ore Deposits. It is a laboratory operated by the School of Earth Sciences at the University of Tasmania
10.	Concentration	The concentration of the element in the sample.
11.	Count per second	This is the number of electrons measured by the detector. Besides, this is the y-axis in the tab-chart.
12.	Cumulative sums of difference (CUSUM)	A method solves the problem of change point analysis.
13.	Damp-trend linear exponential smoothing model	It is a time series model that has two equations (i.e. smooth and trend) and three parameters.
14.	Data reduction	This term means the removal of variation or fluctuation from the tab-chart.
15.	Detection limit	It is the lowest concentration that we can distinguish from the background concentration.
16.	Detector	This is one of main component in ICP-MS instrument. It is responsible for measuring number of electrons from the chosen element.
17.	Drop	It is a significant signal (i.e. count per second) change in the tab-chart. This occurrence means there is a decrease in the amount of the

		element in the sample.
18.	Durbin-Watson statistic	This is a statistical test to check whether the time series is independent or not.
19.	Fitted error	In this thesis, this is the difference of count per second and fitted value (a) or the difference of count per second and fitted value (b).
20.	Fitted curve	This is the curve of fitted value (a) or fitted value (b).
21.	Fitted value (a)	This value is equal to the smooth value of the HOLT model. It is used to fit the tab-chart.
22.	Fitted value (b)	This value is equal to the sum of trend and smooth values of the HOLT model. It is used to fit the tab-chart.
23.	HOLT model	This model is obtained by modifying the linear Holt exponential smoothing model. This modified model has two purposes. Firstly, the fitted curve of this model is used to keep track the trend of the tab-chart. Secondly, the estimates of this model (i.e. trend and smooth equation) are used to develop the method to detect the change points in the tab-chart.
24.	ICP-MS instrument	This is an instrument, which is responsible for analysing the composition of ablated material.
25.	Integration interval	This interval is the starting and ending point of a plateau, which is determined by the visual judgement of the researcher. This interval has consistent trend or slope, assumed to represent homogenous phases or zones in a sample.
26.	Internal standard	It is used to account for the systematic variations in signal intensity unrelated to real change in the composition of the material.
27.	Intuitive method	The division of tab-chart and selection of integration interval are done by the researchers' experience, knowledge and visual judgement.
28.	Jump	It is a significant signal (i.e. count per second) change in the tab-chart. This occurrence means there is an increase in the amount of the element in the sample.
29.	LA-ICP-MS	This term stands for Laser ablation-inductively coupled plasma-mass spectrometry. It is a new technique for measuring the inorganic composition of the minerals or sample
30.	Laser ablation process	A process of using laser to detach material from a sample.
31.	Level	The signal intensity of the element in the background or plateau.
32.	Line analysis	This is one type of laser operation on the sample
33.	Linear Holt exponential smoothing model	It is a time series model that has two equations (i.e. smooth and trend) and two parameters.

34.	Locally weighted regression	This is a non-parametric fitting (smoothing) model. Note that the HOLT model is a parametric model.
35.	Machine drift	ICP-MS instrument will deviate from initial setting and incur error in the measurement. The instrument is calibrated and corrected for every fixed period.
36.	Mass discriminator (spectrometer)	This is one of main component in ICP-MS instrument. It is responsible for selecting the specific ionised species for detection.
37.	Mass number	The weight of the element in a sample.
38.	Mean (average level) of background	This term is the value by averaging the count per second or fitted value (a) or fitted value (b) over the background.
39.	Mean (average level) of plateau	This term is the value by averaging the count per second or fitted value (a) or fitted value (b) over the plateau.
40.	Mean (average level) of the element concentration	This mean is obtained by subtracting mean (average level) of background from mean (average level) of plateau, and then multiply by a constant factor, λ.
41.	Measurement cycle	This is a period which the detector measures a pre-selected list of isotopes (elements).
42.	Microsoft excel	This is a software of arithmetic computation and statistical analysis. The laboratory uses this software to convert the information from ICP-MS instrument into tab-chart.
43.	Misclassification	This term means that the tab-chart can be successfully classified by the classification method or classification rules. In the other words, the method or rules has the worse performance on this tab-chart.
44.	Nu wave UP-213	This machine is responsible for the ablation process. It is equipped with video camera, microscope, motorized stage and laser.
45.	Observation value	This term is equivalent to count per second in this thesis
46.	Ordinary least squares	The formula of calculating the regression line.
47.	Plasma (ICP torch body)	This is one of main component in ICP-MS instrument. It is responsible for creating ionised species.
48.	Plateau	It is a portion of tab-chart, which is used for calculating the concentration of element.
49.	Perceptually important points (PIP)	A method solves the problem of change point analysis.
50.	Primary standard	It is also named standard glass. The two common primary standards are NIST 612 and BCR2. The purposes of these glasses are to correct the drift of the ICP-MS instrument and calculate the elemental concentrations.
51.	R-software	It is a statistical package. This software is designed and specialized in statistical analysis.

52.	Secondary standard	This is another glass standard. It is used to check analytical precision and accuracy during routine LA-ICP-MS runs.
53.	Signal (intensity)	This term is equivalent to count per second in this thesis
54.	Simple exponential smoothing model	It is a time series model that has one equation (i.e. smooth) and a smoothing level parameter.
55.	Solution-based ICP-MS	It is the conventional technique for measuring the inorganic composition of the minerals or sample. The material is digested (dissolved) for analysis.
56.	Spot (vertical) analysis	This is one type of laser operation on the sample
57.	Standard deviation of background	This term is the value by calculating the standard deviation of count per second or fitted value (a) or fitted (b) over the background.
58.	Standard deviation of fitted error	It is obtained by calculating the standard deviation (error) of fitted error over background or over plateau.
59.	Standard deviation of plateau	This term is the value by calculating the standard deviation of count per second or fitted value (a) or fitted (b) over the background.
60.	Standard deviation of the element concentration	This standard deviation is obtained by addition of square of standard deviation of background and square of standard deviation of plateau, and then takes square root and then multiplies a constant factor, λ.
61.	Standard error of background mean	This standard error is calculated by dividing standard deviation of background by square root of number of count per second or fitted value (a) or fitted value (b) over the background.
62.	Standard error of HOLT model over background	This standard error is obtained by calculating the standard deviation of background by using fitted value (a) or fitted value (b) over the background, and then dividing it by square root of number of fitted value (a) or fitted value (b) over the background.
63.	Standard error of HOLT model over plateau	This standard error is obtained by calculating the standard deviation of plateau by using fitted value (a) or fitted value (b) over the plateau, and then dividing it by square root of number of fitted value (a) or fitted value (b) over the plateau.
64.	Standard error of Intuitive method over background	This standard error is obtained by calculating the standard deviation of background by using count per second over the background, and then dividing it by square root of number of count per second over the background.
65.	Standard error of Intuitive method over plateau	This standard error is obtained by calculating the standard deviation of plateau by using count per second over plateau, and then dividing it by square root of number of count per second over the plateau.
66.	Standard error of plateau mean	This standard error is calculated by dividing standard deviation of plateau by square root of number of count per second or fitted value (a) or fitted value (b) over the plateau.

67.	Statistical process control	This is a technique to monitor the process and detect any occurrence of out of control and abnormal pattern.
68.	Statistical summaries	The calculation of mean and standard deviation (error) over background or plateau after the intuitive method, classification method or classification rules is applied to the tab-chart.
69.	Tab-chart	This is a time-series graphical representation of the element distribution in a sample made by the ICP-MS instrument.
70.	Threshold value (T)	This is a parameter of HOLT model. It controls the trend and smooth parameters of the HOLT model.
71.	Time series segmentation	This term is equivalent to change point analysis.
72.	Trend and smooth parameters	These are parameters of HOLT model. They are used to control the weightings on the current observation and the previous observations of the tab-chart.
73.	Trend-smooth plot	The plot of the trend values against smooth values after the HOLT model is well fitted to the tab-chart. It is helpful in developing the rules of dividing the tab-chart into different portions (i.e. classification rules).